Title
Field Implementation of Freeway Control

Author
Wu, Cheng-Ju

Publication Date
2019

Peer reviewed|Thesis/dissertation

Field Implementation of Freeway Control

By

Cheng-Ju Wu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Roberto Horowitz, Chair
Professor Francesco Borrelli
Professor Alexander Skabardonis

Summer 2019

Field Implementation of Freeway Control

Abstract

Field Implementation of Freeway Control

by

Cheng-Ju Wu

Doctor of Philosophy in Engineering - Mechanical Engineering

University of California, Berkeley

Professor Roberto Horowitz, Chair

This dissertation presents a series of traffic management studies including freeway and intersection traffic modeling, estimation, control methodologies and field implementation tests. First, a traffic flow prediction method that combines the most recent traffic data with historical traffic data is studied. An autoregressive moving average with exogenous input (ARMAX) model is estimated on-line based on the most recent vehicle detector station (VDS) data. Results obtained using empirical freeway mainline and on-ramp data show that this method outperforms methods that rely only on the historical average of the data to perform a prediction, especially during days with unusual traffic flow demands.

Second, two freeway control strategies: coordinated ramp metering (CRM) and variable speed advisory (VSA) are investigated and implemented in field tests. In the control of CRM study, the freeway is modeled by the cell transmission model (CTM) and the control problem is solved by the model predictive control (MPC) scheme. The proposed CRM is deployed in a segment of California State Route 99 Northbound (SR-99N) for a five-week field test. The test results shows that the freeway efficiency can be improved by 7.25% for morning peak hours. In the VSA control study, an advisory speed limit control is designed by using traffic flow stabilization of the Lighthill-Whitham-Richards (LWR) model. The proposed VSA is deployed in a segment of California State Route 78 Eastbound (SR-78E) for a five-week field test. The test results shows that the freeway efficiency can be improved by 8.71% for morning peak hours. Both control strategies indicate freeway efficiency improvement in congested traffic.

Third, the large-scale signalized intersection traffic network control by offset optimization is also studied. The traffic network is described by a directed graph and the traffic dynamic is represented by continuous-time fluid queue model with sinusoidal arrival and departure rate assumption. The original non-convex offset optimization problem can be relaxed into a semidefinite program (SDP). The Burer-Monteiro (BM) method is used for solving the large SDP to avoid conic constraints. Two real-world traffic simulation networks, respectively in Manhattan, NY and in Pasadena, CA are constructed for demonstrating the BM

method. Numerical simulation results indicate that BM method has good scalability and it can efficiently recover optimal solutions of the SDP.

To my family

Contents

List of Figures

List of Tables

Acknowledgments

I have appreciated the learning experience at UC Berkeley throughout my Ph.D. career. It has been a challenging but at the same time rewarding experience. I greatly appreciate the support and guidance from my research advisor and the chair of this dissertation committee Professor Roberto Horowitz, especially during the toughest times of my academic career.

I would like to thank Professor Kameshwar Poolla, Professor Francesco Borrelli, Professor David Brillinger and Professor Lin Lin, for serving in my qualifying exam committee. I would also like to extend my thanks to the dissertation committee members Professor Francesco Borrelli and Professor Alexander Skabardonis.

I would like to thank Thomas Schreiter and Gabriel Gomes for their collaboration on traffic flow estimation described in Chapter 3.

I would like to thank my Richmond Field Station colleagues, Dr. Xiao-Yun Lu, Dr. Steven Shladover, and John Spring for their collaboration and technical guidance throughout the freeway control field test implementation project described in Chapter 4, 5 and 6.

I would like to thank Eric Kim and Professor Murat Arcak for their collaboration on signalized intersection offset optimization described in Chapter 7.

I would like to thank my parents, Jan-Hsun Wu and Shu-Nuan Hsieh, for their support and sacrifices to raise me and educate me. I could not have been where I am today without their help and guidance.

Chapter 1

Introduction

1.1 Background

Due to world population growth and global urbanization trends, traffic congestion has become a worldwide issue that has a significant detrimental impact on human society's productivity. Transportation agencies are not able to meet the raising demand, especially during commuting peak hours, due to inadequate infrastructure, lack of funds or space for infrastructure expansion, or no proper traffic management. In these circumstances, the inevitable result is congestion in metropolitan areas, urban areas or the key transit roads (freeway, intersections, or main street). Both of recurrent and nonrecurrent traffic congestion causes various of problems such as travel delays, reducing capacity of traffic infrastructures, waste of fuel, air pollution, higher accident risk, and other potentially unsafe conditions. According to the 2015 Urban Mobility Scorecard [94] presented by the Texas A&M Transportation Institute, the national congestion cost in 471 U.S. urban areas was $160 billion in 2014, $114 billion in 2000 and $42 billion in 1982, however, the cost will grow from $160 billion to $192 billion in 2020 (in 2014 dollars), which corresponds to the average commuters congestion yearly cost will grow to approximately $1,100 in 2020 (in 2014 dollars). In addition, delay will grow to 8.3 billion hours in 2020 and wasted fuel will increase to 3.8 billion gallons in 2020. As a consequence, the average commuter will waste 47 hours and 21 gallons in 2020 due to congestion. Considering the increasing demand, growth in car ownership and slow infrastructure improvements, it can be conjectured that the congestion problem is bound to continue.

Building new transportation infrastructure to alleviate congestion and accommodate higher traffic demand is not always the best solution because of negative environmental impact, economical concerns, political reasons or many other limitations. Constructing more roads or widening existing ones will require more fundamental land use and the road expansion costs will probably be high. Besides, such solutions may not be feasible in metropolitan areas because of land shortage. Even when roads are expanded and congestion is resolved, population could move from congested areas to less congested areas, and therefore traffic

demand for the latter will increase, which will in turn cause congestion again. Instead of constructing new infrastructure, developing traffic management techniques, which can make a better use of existing traffic infrastructure, has been proposed and partially tested in field implementation during the past few decades, to reduce traffic congestion. If road expansion is not able to eliminate traffic congestion, introducing proper traffic management technology are necessary to maintain road network mobility and safety. Smart cities in the future will require Intelligent Transportation Systems (ITS) that mitigate congestion problems, and the next generation of transportation systems will include connected infrastructure, connected vehicles and automatic vehicles.

The general procedures of constructing a traffic management system from design to implementation is suggested as following: 1) collect traffic data from the field; 2) analysis and understand the traffic dynamic characteristics by the collected traffic data; 3) build traffic models and calibrate the model; 4) design traffic control algorithms and evaluate the control performance by simulation; 5) implement control methodologies in the field using available hardware devices; and 6) evaluate the field implementation results by traffic performance indices.

Collecting traffic data from the field is the first step of studying traffic characteristics. Traffic data can be obtained by surveys, manual collection, or detection with sensors, depending on the requirement and availability of collection techniques. Vehicle detection technologies may be classified as non-intrusive or intrusive, depending on whether they need to be embedded or installed in the road pavements. Non-intrusive detection technologies include video cameras, microwave radars, light barriers, and pneumatic tubes, which are often susceptible to weather disruptions and vandalism. Intrusive technologies include inductive loop detectors, magnetometers, pressure cell, strain gauge, and fiber-optic sensors, which can often be rather costly for installation and suffer from a low survival rate. Moreover, it is a very challenging and costly task to maintain the sensor network and to manage the huge amounts of data generated by a dense array of wired sensors. If the desired data cannot be directed measured from the field or the data quality is not acceptable, developing feasible and reliable estimation methods is necessary.

Understanding traffic characteristics and dynamics is important for solving congestion problem because it serves to identify the cause of congestion and to predict the influence of traffic control. Recurrent or non-recurrent traffic pattern can be recognized by studying historical traffic data, which provides traffic modeling guidelines to researchers. The identified real-world traffic phenomenon is usually described by a set of mathematical equations, which is called traffic model. A well-calibrated traffic model are utilized to evaluate the proposed control method. Many traffic models represent a specific traffic phenomenon, but there is no generic traffic model that can represent all traffic phenomena. Therefore, traffic control designers have to properly select models that precisely represent the traffic phenomena of interest and fulfill the control requirements.

Traffic control is an important operational management strategy that can be used to alleviating traffic congestion. For freeway control, ramp metering (RM) and variable speed advisory/limits (VSA/VSL) are two commonly used strategies to regulate traffic flow and

delay the onset of congestion. For intersection control, the most common strategy is fixed-time signal controller, which is governed by three control parameters: cycle length, green splits, and offset. A traffic control strategy will improve the network performance only when it is properly designed. In practice, traffic control designers not only need to test the proposed strategies by the simulation model, but also need to overcome the implementation limitation: traffic regulations, driver's acceptance, hardware and software restriction and accessibility of required infrastructure.

1.2 Dissertation Organization

This dissertation is organized as follows. In Chapter 2, we review previous research related to the topics discussed in this dissertation. These include traffic data collection technologies, freeway traffic models, and commonly used freeway traffic management methodologies.

In Chapter 3, a novel freeway on-ramp flow demand prediction technique, which combines the strengths of a model that uses a large amount of historical data with an autoregressive moving average with exogenous input (ARMAX) model with fast computation, is proposed. In the proposed traffic flow prediction method, historical data are aggregated to flow profiles, which represent a meaningful average for each weekday. To reduce the effect of noise and obtain an accurate historical average, multiple months of data are aggregated to seven typical daily profiles for each detector. Then, flow profiles that represent a typical day, one for each day of the week is calculated. Since this computation involves a large amount of data, it can be computationally demanding. This, however is not a problem in practice, since these nominal historical profiles can be computed in an off-line fashion, so that no running time constraints apply. Using the nominal historical profile as a deterministic input and the actual traffic flow as the noise-contaminated measured output, we identify in a real-time recursive fashion the parameters of an autoregressive moving average with exogenous input (ARMAX) model that best describes this input/output relation. Using the results of the ARMAX identification process, the optimal multi-step ahead predictor model of the traffic flow, which utilizes the flow measurements up to the current time and the nominal historical profile by solving a Bezout equation can be determined. Due to its relatively low on-line computation time, this system can be applied in practice to predict traffic flow in real-time. The relevant publication for this chapter is listed below.

- **Cheng-Ju Wu**, Thomas Schreiter, Roberto Horowitz, and Gabriel Gomes (2014), *"Traffic Flow Prediction Using Optimal Autoregressive Moving Average with Exogenous Input-Based Predictors"*, In Transportation Research Record: Journal of the Transportation Research Board, Volume 2421, Issue 1, Pages 125-132.

In Chapter 4, a procedure for building a microscopic freeway traffic model and calibrating it is demonstrated. First, a brief description of the studied network SR-99N is presented, which includes the network configuration and the deployment of vehicle detector

station (VDS). Second, the traffic demand estimation and split ratio calculation are presented. Third, the traffic analysis for SR-99N is discussed. Then, the model calibration method is explained, which includes how the model parameters are adjusted and the microsimulation results are demonstrated. Although this study simulates a single site using a specific commercially-available modeling software (AIMSUN), practitioners who want to use microsimulation to model a freeway with multiple vehicle class and HOV lane will find details on how to build a freeway traffic model using a microsimulation software, and how to overcome problems associated with incomplete and missing data. The relevant publication for this chapter is listed below.

- **Cheng-Ju Wu**, Xiao-Yun Lu, Roberto Horowitz, and Steven E Shladover (2016), *"Microsimulation of Congested Freeway with Multiple Vehicle Classes Using AIMSUN: A Case Study on SR-99N"*, In Transportation Research Board 95th Annual Meeting.

In Chapter 5, a field test implementation of a coordinated ramp metering (CRM) system is presented. The freeway traffic dynamics is modeled via the cell transmission model (CTM). By introducing traffic performance indexes as the objective function, the control problem is formulated under the framework of optimal control and the ramp metering rate is calculated by a model predictive control (MPC) scheme. The system was implemented during a five-week period from October 3 to November 4, 2016 on a test segment of SR-99 Northbound, which is one of Sacramento's most congested freeway. The field test effectiveness in reducing traffic congestion and improving freeway efficiency was evaluated using PeMS data, which was independent of the data used by the CRM system for objectivity in the performance evaluation. The relevant publication for this chapter is listed below.

- **Cheng-Ju Wu**, Xiao-Yun Lu, John Spring, and Roberto Horowitz (2018), *"Field Test Implementation of Coordinated Ramp Metering Control Strategy: A Case Study on SR-99N"*, In Transportation Research Board 97th Annual Meeting.

In Chapter 6, a field test implementation of a variable speed advisory (VSA) system is presented. The proposed VSA algorithm adopts occupancy and speed measurement from upstream and downstream of a VSA location, and does not employ traffic models in calculating the time-variant advisory speed values. By providing advisory speed to the drivers on the freeway, the system is able to improve the capacity flow of a bottleneck located downstream of the VSA station and reduce speed variations and the potential for rear-end collisions upstream of the bottlenecks. The VSA system was field tested during a four-week period from April 9 to May 4, 2018 on a test segment of SR-78 Eastbound, which is one of San Diego's most congested interchanges. The field test effectiveness in reducing traffic congestion and improving freeway efficiency was evaluated using PeMS data which is independent of the data used by the VSA system for objectivity in the performance evaluation. The relevant publication for this chapter is listed below.

- **Cheng-Ju Wu**, Xiao-Yun Lu, John Spring, and Roberto Horowitz (2019), *"Field Test Implementation of Variable Speed Advisory Control Strategy: A Case Study on SR-78E"*, In Transportation Research Board 98th Annual Meeting.

In Chapter 7, the offset optimization of a traffic light signalization in a large-scale arterials traffic network is studied. The traffic network is described by a directed graph and the traffic dynamic is represented by continuous-time fluid queue model under a sinusoidal arrival and departure traffic flow rate assumption. The resulting non-convex offset optimization problem can be relaxed into a semidefinite program (SDP) and then the large SDP problem is solved by the Burer-Monteiro (BM) method to avoid conic constraints. Two real-world traffic networks, respectively located in Manhattan, NY and in Pasadena, CA, were constructed for demonstrating the BM method. Numerical simulation results indicate that BM method has good scalability and it can efficiently recover the optimal solutions of the SDP. The relevant publication for this chapter is listed below.

- Eric S. Kim, **Cheng-Ju Wu**, Roberto Horowitz, Murat Arcak (2017), *"Offset optimization of signalized intersections via the Burer-Monteiro method"*, In 2017 American Control Conference (ACC).

Finally, Chapter 8 summarizes the work presented in this dissertation and discussed possible directions for future research.

Chapter 2

Review of Related Works

2.1 Traffic Data Collection

Traffic measurements and data archival are essential components of any intelligent transportation system designed for traffic networks management. Traffic engineers usually use historical data to predict future traffic patterns, plan new infrastructures, calibrate traffic models, and test control strategies. Traffic engineers also used real time traffic data in the day to day operations to implement traffic-responsive control, manage incident/accident, and provide travel information to drivers.

Traffic Measurements

The most commonly obtained traffic measurements are occupancy and vehicle count. From the occupancy and vehicle count measurements, other macroscopic quantities, such as vehicle density, speed and flow, can be derived. Traffic quantities that are commonly used in describing traffic states are listed below.

Occupancy Occupancy is the percentage of time when a traffic sensor, for example, a loop detector, is occupied by vehicles, i.e., the fraction of time when there are vehicles inside the detection zone of that sensor.

Vehicle count (volume) Vehicle count is the total number of vehicles recorded by a sensor that pass over its detection zone within a given time interval.

Speed Vehicle speed can be aggregated over space or time. Time-mean speed is calculated by averaging vehicle speeds over time at a fixed position, while space-mean speed is calculated by averaging vehicle speeds over a space interval at a fixed time. Vehicle speed measured by stationary sensors is the time-mean speed, or called point speed.

Flow Flow is the number of vehicles passing a fixed position per unit time .

Density Density is the number of vehicles per unit length at a fixed time.

Traffic Detection Technologies

Modern innovations are bringing new functionalities to traffic detection [68] [49]. Various vehicle sensory technologies are deployed in the traffic network to measure traffic data. Traffic detection technologies can be classified as intrusive and non-intrusive type, depending on whether the sensor need to be embedded in the road pavement. Intrusive sensors includes inductive loop, magnetometer and pneumatic tubes. Non-intrusive sensors include video detection systems, microwave radars, and probe-based traffic monitoring system. The main characteristics of these detection technologies are listed below.

Inductive loops Inductive loops have been widely used in traffic flow and occupancy detection since the 1960s. Inductive loops detect induced eddy current when vehicles presence on top of the detectors. Speed information can be obtained if dual loop is available. Their detection accuracy is high, but their deployment causes significant traffic disruption and they are highly susceptible to malfunction since they have to be embedded in the pavement for installation.

Magnetometers Magnetometers detect vehicles presence on top of the detector using the changes in magnetic fields. Comparing with loop detectors, the installation of a magnetometer is easier and its size is smaller. Besides, it usually provides wireless communication to the signal controller.

Pneumatic tubes Pneumatic tubes detect vehicle axles when tires run over the tubes. These are deployed perpendicular to the road surface. The sensor is non-intrusive, so it can be quickly installed and mostly deployed for short term studies.

Video detection systems Video detection systems uses cameras to record images and convert vehicle presence and movement data by applying an image processing algorithm. These can be used for detecting road incident, tracking vehicle trajectories, vehicle types counts, and vehicle speed. Their main advantage is that a single camera can monitor multiple lanes and zones. The data quality is affected by weather, shadows and vehicle occlusion, so cameras need regular maintenance.

Microwave radars Microwave radar send energy towards an area of the roadway and receive reflected energy to detect vehicle presence and track vehicle movement, which is the application of Doppler effect. They are generally insensitive to weather conditions, but Doppler sensors cannot detect stopped vehicles.

Probe-based traffic monitoring system Drivers share their global positioning system (GPS) data on the mobile phones in real-time during most of their drives. These position information is transmitted to a receiver (data collection center) to display real-time trajectory of the probe vehicles. The collected data is usually called probe data (floating car data), which can be used to estimate travel time information, speed and origin-destination demand matrices.

The Freeway Performance Measurement System

The Freeway Performance Measurement System (PeMS) is a traffic data archival system, which collects data in real-time from nearly 40,000 individual detectors (as of 2019) spanning the freeway system across all major metropolitan areas of California [16] [86]. As of 2019, the data sources in PeMS covered nine Caltrans Districts out of 12, specifically, district 3 (North Central), 4 (Bay Area), 5 (Central Coast), 6 (South Central), 7 (Los Angeles and Ventura counties), 8 (San Bernardino and Riverside), 10 (Central), 11 (San Diego and Imperial), and 12 (Orange County). The collected data is filtered, processed and stored in databases. Users can access PeMS over the Internet through a Web browser, where provides both historical and real-time performance measurements of the freeways. PeMS has been extensively used by researchers, traffic engineers, planners, traveler information services over its years of operation since 1998.

PeMS receives 30-second loop detector data (flow and occupancy) in real time from each Caltrans District Transportation Management Center (TMC) to produce useful information. In the case of dual loops, PeMS also receives speed data. Most loop detectors in California have single loops. PeMS uses an adaptive algorithm to calculate the g-factor (effective vehicle length) from the flow and occupancy data and provides accurate speed estimation for single loop detectors [42]. PeMS aggregates the processed data into 5-minute intervals, which fills in the missing data with plausible values by imputation algorithms. PeMS also has detailed diagnostic measures to ascertain the operation of the detector and the quality of the data. PeMS uses collected data to compute basic freeway performance measurements, including Vehicle Miles Traveled (VMT), Vehicle Hours Traveled (VHT), Congestion Delay, and the freeway efficiency "Q" (ratio of VMT over VHT) interpreted as the average speed in the region of analysis.

Next Generation Simulation

The Next Generation Simulation (NGSIM) program [74] is a project supported by United States Department of Transportation (US DOT) Federal Highway Administration (FHWA). The goal of this program is to develop a core of open behavioral algorithms for traffic simulation with a primary focus on microscopic models. The NGSIM program provides vehicle trajectory data of high resolution and detailed supporting documents, which are open and free for the simulation community. The traffic data collected by the NGSIM program contains typical free-flow and congested traffic. Four data-sets collected from freeway segments and surface street are provided by this program. The freeway data was collected from eastbound I-80 in Emeryville, CA and southbound US-101 in Los Angeles, CA; the surface street data was collected from Lankershim Boulevard in Los Angeles, CA, and Peachtree Street, Atlanta, GA. The two freeway data-sets are 45 minutes long each and the two arterial data sets are 30 minutes long each. The vehicle trajectory data of NGSIM was collected through a network of synchronized digital video cameras, which capture individual vehicle's position every 0.1 second and it provided the precise location of each vehicle within the study area.

Because the data contains detailed vehicle trajectories, it can be used in studying driver behavior, especially for microscopic modeling.

2.2 Freeway Traffic Modeling

There are three different approaches to model traffic: microscopic, mesoscopic, and macroscopic [41]. In microscopic models, vehicles are modeled as discrete entities that can have different properties, while in macroscopic models all the vehicle properties are lumped into continuous variables, such as density, speed and flow, of the traffic stream depending on time and position on the road. In mesoscopic models, traffic is considered by groups of objects, and the activities of groups and interactions among groups are described at a lower level. The microscopic approach seeks to reproduce the behavior of the individual vehicle/driver entities, such as lane changing, car following, and their other responses to the driving environment by adjusting its speeds. The macroscopic approach ignores the dynamics of the individual vehicle/driver and instead attempts to replicate the aggregated response of a large number of vehicles. The mesoscopic approach does not distinguish nor trace individual vehicles, but specifies the behavior of individuals in probabilistic or statistical ways.

Macroscopic Traffic Flow Models

Macroscopic models describe the evolution of aggregated traffic behaviors: density, speed and flow over space and time using a set of partial differential equations (PDEs), together with other constituent relationships. Each model has a basic conservation law, which captures the fact that vehicles cannot be created or destroyed. Well-known examples of macroscopic traffic flow models are Lighthill-Whitham-Richards model (LWR model) [59] [91], Cell Transmission Model (CTM) [23], Asymmetric Cell Transmission Model (ACTM) [30] and the METANET model [51]. Important models related to this dissertation are reviewed in this section.

The Lighthill-Whitham-Richards Model

The Lighthill-Whitham-Richards partial differential equation, also known as the LWR model, is a first order model described by the vehicle conservation equation

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial f(x,t)}{\partial x} = 0$$

and the static flow-density relationship

$$f(x,t) = \rho(x,t)v(x,t) = \Phi(\rho(x,t)) \tag{2.1}$$

where $\rho(x,t)$ is density at location x and time t, $f(x,t)$ is flow, $v(x,t)$ is speed, and $\Phi(\rho)$ is flux function. The curve $\Phi(\rho)$ describe the relation between flow and density and it is also called the fundamental diagram in traffic engineering [34]. Later researchers have suggested

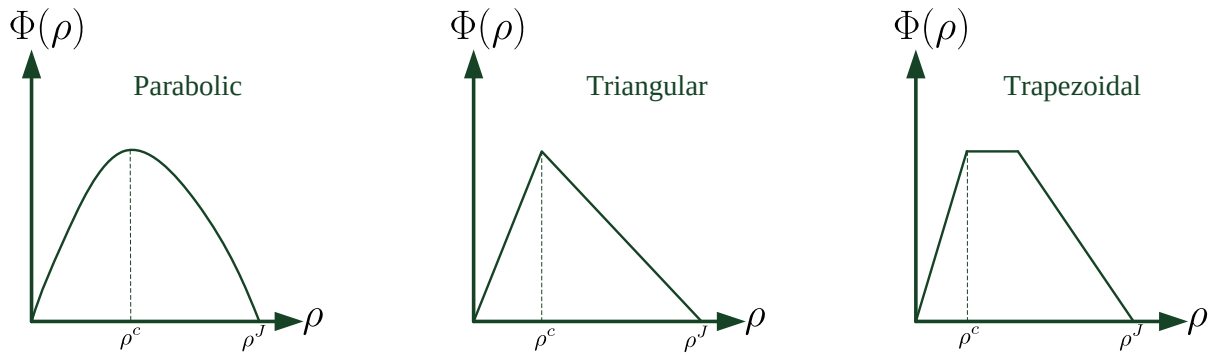


Figure 2.1: Commonly used fundamental diagrams (flux function $\Phi(\rho)$)

various shapes of fundamental diagrams that provide a better fit to the measured data [24]. Figure 2.1 shows some of the commonly used fundamental diagrams. In general, the fundamental diagrams has the following properties.

1. $\Phi(\rho) \geq 0$ is a concave continuous function.
2. $\Phi(0) = \Phi(\rho^J) = 0$, where ρ^J is the maximum density in the road which is known as the jam density.
3. $\Phi(\rho)$ attains a maximum flow at ρ^c, which is the critical density. The maximal flow is call the capacity.

The critical density separates the fundamental diagram into two regions: the free flow region when $\rho \leq \rho^c$ and the congested region when $\rho > \rho^c$. If traffic state is in free-flow region, vehicle speed is dependent on the posted speed limit of a road or the desired speed of a driver, but not restricted by other vehicles. The speed in this regime is called the free-flow speed. On the other hand, the speed of a vehicle is restricted by surrounding vehicles if it is traveling in the congested regime.

The Cell Transmission Model

The Cell Transmission Model (CTM) proposed by Daganzo [23] is a first order discrete model, which is the discretization of the LWR model with triangular fundamental diagram. Under the CTM framework, the highway is divided into homogeneous, consecutively numbered sections of length L_i, where i is a section index. Time is discretized into uniform intervals of duration T_s, such that

$$v_f T_s \leq \min_i L_i \tag{2.2}$$

where v_f is the free-flow speed. The uniform sections are known as cells, and they are increasingly numbered from upstream to downstream. The density of cell i can be represented

by the conservation equation

$$n_i(k+1) = n_i(k) + f_{i-1}(k) - f_i(k) \tag{2.3}$$

where $n_i(k)$ is the number of vehicles in cell i at time step k, and $f_i(k)$ is the flow moving from cell i to cell $i+1$ during the time step k. The flow in CTM is obtained by comparing the sending and receiving flows of a cell, also known as the demand and supply, as

$$\begin{aligned} f_i(k) &= \min(D_i, S_{i+1}(k)) \\ D_i(k) &= \min(F_i, n_i(k)V_i) \\ S_i(k) &= \min(F_i, W_i(n_i^J - n_i(k))) \end{aligned} \tag{2.4}$$

The fundamental diagram parameters can be different for each cell, and the parameters are indexed by the cell number i. The triangle fundamental diagram is characterized by the free-flow speed V_i (the slope of the free-flow region of the fundamental diagram), the maximum flow F_i (the capacity), the congestion wave speed W_i (the slope of the congested region) and the jam density n_i^J (the maximal density). The demand function $D_i(k)$ describes the flow that can be sent from the upstream cell to the downstream cell, while the supply function $S_{i+1}(k)$ specifies the maximum flow that can be received by the downstream section. The flow is obtained by taking the minimum of the demand and the supply.

Microscopic Traffic Flow Models

With the increasing availability of very fast computers and concurrent progress in the development and understanding of efficient algorithms, microscopic simulation model have been widely used in both transportation operations and management analyses. These models distinguish and trace individual vehicle/driver and the driver's behavior is generally described by a large set of logical rules. By selecting proper driver individual and vehicle characteristics, position, speed and acceleration of each vehicle are calculated for each simulation time step. The time step is usually chosen in the range of 0.1-1.5 seconds, depending on simulation accuracy. Each vehicle trajectory in the microscopic model are usually generated by car-following model, lane-changing model, and gap-acceptance model.

The car-following model describes the response of vehicles to other vehicles around them. It determines the speed that drivers maintain under a certain traffic condition, the headway or spacing between vehicles, and the acceleration or deceleration of vehicles when drivers need to adjust their speeds. The lane-changing model defines the longitudinal behavior of vehicles, which contains change lanes, merge, and weave. Reasons to change lanes when driving are generally fall into the following two categories: mandatory lane change (changing lanes to get into the correct lane for the destination) and discretionary lane change (changing lanes to get gain in speed). Once drivers decide to make a lane change, they have to judge whether it is possible to change lanes and plan the actions to take in order to ensure a safe lane change. The gap-acceptance model demonstrates the behavior when vehicles turn into

or across conflicting traffic flows. It defines the conditions when a gap is acceptable for a safe turn or crossing.

Microscopic models often utilize a set of parameters to specify drivers' behavior. Some of these parameters cannot be directly measured by loop detector from the field or it is hard to measure. For instance, reaction time, which accounts for the time a driver takes to perceive, recognize, and respond to a road event, is used in car-following models. Microscopic simulation tools usually allow users to select different parameter values for different vehicle types, and the model parameter is in a probability distributions. Therefore, calibrating model parameters is usually a time-consuming task in microscopic simulation. Besides, the computation time of executing microscopic simulation is longer than macroscopic simulation because the state of each vehicle is updated in every simulation time step. Examples of microsimulation model software packages are AIMSUN [2], VISSIM [106], CORSIM [104], and PARAMICS [82]. These software programs differ in their driver behavior models and modeling capability. AIMSUN is used in several experiments in this dissertation.

2.3 Freeway Traffic Control

Freeway traffic management and operations consist in the implementation of policies, strategies and technologies to improve freeway performance. The main objectives of freeway control include minimizing congestion (and its side effects), improving safety, and enhancing overall mobility [112]. This section gives a brief review of the most common used freeway traffic control strategies: ramp metering and variable speed advisory [78].

Ramp Metering

Ramp metering systems are primary devices for addressing freeway recurrent congestion. Ramp meters are traffic signals placed near the entrance of the freeway on-ramps. When the ramp metering is active, most traffic lights in ramps allow 1 or 2 cars per green per lane. Therefore, regulating the frequency of green lights controls the rate at which vehicles enter the mainline such that the downstream capacity is not exceeded, thereby maximizing the freeway's throughput at a uniform speed. Ramp metering control can reduce travel time by avoiding the mainline capacity drop [116] and offramp blockage [13], and by encouraging certain diversions [115]. In addition, ramp metering has the ability to break up platoons of vehicles that have been released from a nearby-signalized intersection such that the collision probability of freeway merge area can be reduced. Ramp metering could also provide incentives for carpooling and bus ridership by allowing HOVs (High Occupancy Vehicles) to bypass the ramp meter.

Ramp metering can also be classified as traffic responsive, or fixed time, depending on whether the metering actions is activated based on real-time traffic conditions (volume, occupancy, or speed) or not. Ramp metering algorithms can also be classified depending on the scope of their action as either local or coordinated. In local ramp metering control,

the controller only uses the traffic information near a single on-ramp and each adjacent controller are independent and not connected. In coordinated ramp metering controller, controllers share traffic information with each other and several ramp meters coordinate their actions to regulate traffic simultaneously.

Over the years, a number of ramp metering strategies have been developed and deployed various parts of the world. The simplest fixed time controllers activating by time-of-day or day-of-week which specify a fixed metering rate determined from historical data. The most popular local traffic responsive ramp metering algorithm is ALINEA [80] and its variations [98]. Examples of coordinated ramp metering strategies include SWARM (System Wide Adaptive Ramp Metering) [77], HERO (HEuristic Ramp metering cOordination) [27], and METALINE [79].

Variable Speed Advisory

Variable speed advisory (VSA) are enacted by message signs that can be changed to alert drivers when traffic congestion is imminent or incident happening downstream. Traffic sensors along the roadway detect when congestion, incident or bad weather conditions exceed specified thresholds and automatically activate the sign with the message of reducing the speed limit to slow traffic thereby postpone the onset of congestion. The purpose of this mainline speed control is to slow traffic uniformly in a way that creates homogeneous flow, reduces lane-changes and avoids stop-and-go conditions. Depending on the regulations, the posted speeds are advisory in some area, while many require mandatory compliance with enforcement and it is called variable speed limit (VSL). Ideally, the speed limit and message alerts are automated and do not require intervention from any operator. The speed limits change in increments of 5 or 10 mph to progressively regulate the mainline traffic flow.

By traffic management purposes, VSA/VSL systems can be categorized into two main types, homogenization systems and incident detection systems. The main difference between the systems is that homogenization systems are often activated before congested traffic states are reached, while incident detection systems are triggered when a breakdown, i.e. situations with very low speeds, is detected. Therefore, the speed limit of the homogenization systems is often reduced gradually, while the speed limit of incident detection systems have more abrupt changes. Some research showed that incident detection based VSA can reduce the number of incidents [101] [58] and decreased variance mean speed between lanes [75]. Homogenization VSL systems have been shown to improve traffic efficiency [70] [17] and decrease time needed to resolving moving shockwaves [38].

Comparing with ramp metering, VSL has more advantage of preventing capacity drop since VSL can be directly applied to the upstream area of the bottleneck location and drives can receive the control action immediately by message without delay. In contrast, if the upstream on-ramp is far away from the bottleneck location, the control effect of mainline flow will be a large time delay when ramp metering based controllers are used. In more complex freeway configuration, if there is an off-ramp in between the on-ramp and the bottleneck location, the effectiveness of ramp metering may be limited.

Chapter 3

Traffic Demand Prediction

3.1 Introduction

Traffic Management Centers (TMC) want to improve the performance of road networks and reduce congestion by actively managing the infrastructure of a freeway corridor. A promising avenue for proactive traffic management is the prediction of the near-future traffic conditions in real-time by employing a traffic flow model. An important set of calibration parameters of such a model are the boundary flows, i.e., the amount of traffic that is expected to enter the network during the prediction horizon.

Reducing congestion both under recurrent and non-recurrent conditions is a challenge for freeway traffic management centers (TMC). The Connected Corridors program at the University of California PATH program [20] is currently developing Decision Support Systems (DSS) that are able to forecast future short-term traffic conditions and evaluate potential traffic management strategies to improve mobility and safety in freeway corridors. A schematic block diagram of the traffic flow prediction module of a DSS is depicted in Figure 3.1. Many traffic predictors, like [95], use a traffic flow model based on the Cell Transmission Model [23] to predict near-future traffic conditions in real-time, usually with a prediction horizon of approximately one hour. In order to accurately estimate the potential benefits of traffic management strategies such as ramp metering, traffic routing and detouring, and lane management, the traffic flow model has to be calibrated well so that its predictions closely match actual future traffic conditions. Accurate calibration of traffic flow models is still a major challenge and usually requires careful estimation of its parameter values. As Figure 3.1 shows, there are many parameters that affect the prediction, such as network modeling, the fundamental diagrams, estimation of the current traffic state, and the prediction of boundary flows and split ratios. This research focuses on one set of these parameters, namely the prediction of the incoming flows at the boundaries of the road network.

Much research has been devoted to the prediction of traffic flow. The area of traffic investigation known as OD prediction aims at calculating the amount of traffic that is expected to travel between each origin (O) and each destination (D) of the road network. Examples

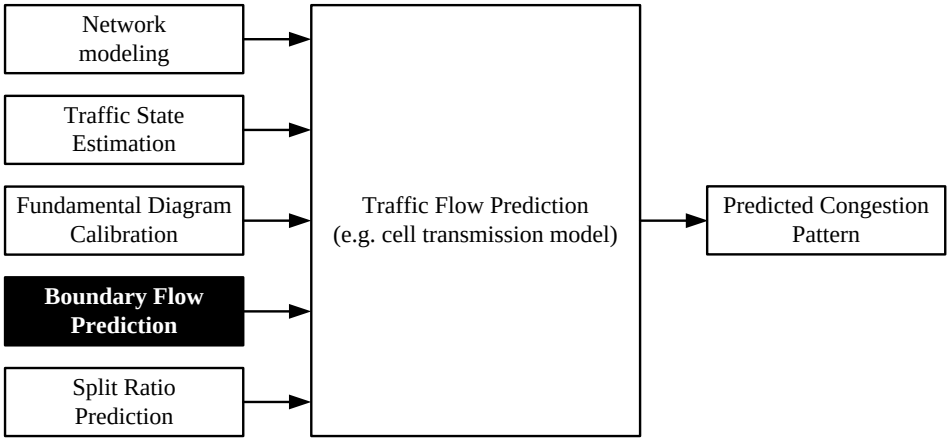


Figure 3.1: Components for on-line traffic flow prediction

include [4], [118] and [25]. An advantage of OD prediction is that the resulting flow information is detailed and permits the search for true optimal traffic management strategies. However, OD estimation is a difficult problem, which is due to the fact that the available loop detector data is generally insufficient to determine the OD flow uniquely. As a consequence, the implementation of optimal traffic management strategies based on OD predictions may not yield the expected performance gains if the true OD patterns cannot be determined. Employing Lagrangian mobile data may provide partial flow information for each OD pair, which may significantly improve the quality of OD prediction in the future.

An alternative and somewhat coarser form of boundary flow prediction is to disregard the destination and to only estimate the *total* flow at each origin. (To complete the traffic assignment, one must also predict split fractions at each bifurcation of the network, which determine how much traffic is leaving the network at each off-ramp. Although similar in nature to boundary flow prediction, split ratio prediction is out of the scope of this research.) As shown in [72] and [71], this information can be determined uniquely in freeways, using only historical mainline data flows. In this research we will assume that historical freeway mainline and ramp traffic flow data is available through data collection and repository sites such as the Performance Measurement System (PeMS) [86]. In the following, we briefly discuss existing boundary flows forecasting methods.

One naive forecasting methodology is the zero-order-hold predictor, which applies the latest flow measurement over the prediction horizon. Another simple method is to directly apply a historical average, such as the mean flow of the previous weeks for the same day and time of day.

A more sophisticated forecasting technique is based on autoregressive moving average (ARMA) models, which treat data as a time series and linearly combine previous flow measurements to predict the future flow. The forecasting technique proposed in [54] employs an ARMA model with an integral action (so-called ARIMA) to forecast flow. Other ARMA-like models include a seasonal term (called seasonal ARIMA, or SARIMA), which exploit the

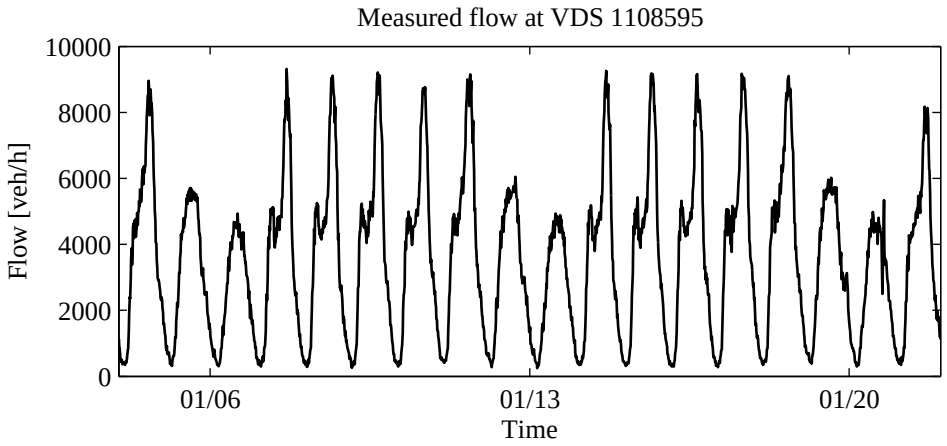


Figure 3.2: Flow measurements over the course of multiple weeks. A daily and a weekly period is clearly visible. (Data from mainline detector station 1108595 between 4 January and 21 January 2013, data source: [86])

periodicity of traffic flow. Figure 3.2 shows the flow profile of a detector over three weeks. One can clearly identify two periods of repeating traffic patterns. Firstly, traffic is following a daily pattern, with high traffic during the day and low traffic during the night. Working days thereby tend to have more traffic and show two peak periods, whereby weekend days exhibit a lower flow with only one peak period. Secondly, a weekly pattern is visible repeating the pattern of five working days and two weekend days.

In many works, the flow of the previous week is combined with recent measurements, for example [111], [96] and [99]. These approaches assume that data from different locations are uncorrelated. To model spatio-temporal correlations between nearby locations, [45] and [44] proposed a spatio-temporal seasonal ARIMA model. Another forecasting technique originates from machine learning. Artificial neural networks were used to forecast boundary flows to capture nonlinear traffic flow phenomena in [107] and [43]. Support vector machines (SVM) are another machine learning technique employed to forecast boundary flows, such as by [117].

Since the literature of traffic flow forecasting is large, several authors have compared different approaches. The research in [47] compared statistical approaches such as ARMA-like models with machine learning approaches and provided many references in transportation beyond the forecast of flow. Recently, [60] compared many approaches experimentally, including simple forecasting methodologies, ARMA-like models, and machine learning models. They confirmed that models that combine historical data with recent data clearly outperform approaches that use only one or the other. For a model to be suitable for on-line traffic prediction applications, it must be both computationally efficient and accurate in its prediction. Simple non-model based prediction approaches are fast, but usually lack the accuracy of more sophisticated model-based or machine learning-based models. On the other

hand, machine learning techniques can be slow and hard to calibrate. A special case is the seasonal ARIMA model: some authors use data only from one previous period (commonly the previous week), which leads to fast computation and good accuracy. The drawback is that if the previous week contained irregular data, for example caused by an extraordinary event such as an accident, the accuracy of the forecast can diminish significantly.

In this chapter, we propose a boundary flow prediction method that combines the most recent traffic data with historical traffic data. An autoregressive moving average with exogenous input (ARMAX) is estimated on-line based on the most recent vehicle detector station (VDS) data. An optimal multiple step ahead traffic demand predictor is obtained based on the estimated ARMAX model by solving a corresponding Bezout equation for each predictor. Results obtained using empirical freeway mainline and on-ramp data show that this method outperforms methods that rely only on the historical average of the data to perform a prediction, especially during days with unusual traffic flow demands, such as a Super Bowl Sunday. Due to its simplicity and robustness, this method is useful in practical applications.

3.2 ARMAX Model

In this section, we present how to determine the multi-step ahead optimal predictor of traffic flow, based on both historical and the most recent traffic data. The concept of flow prediction is illustrated in Figure 3.3 for a vehicle detector station (VDS). In this figure, profile $u(k)$ is the nominal historical flow data for one day, sampled with sampling rate Δt so that the profile consists of $N = \frac{24\text{h}}{\Delta t}$ data points. Profile $y(k)$ is the recent flow data of the prediction day from midnight to the current time step L. Given $u(k)$ and $y(k)$, the proposed prediction algorithm computes the multi-step traffic flow prediction $y^{\text{p}}(k+D|k)$ for a given prediction horizon N_{p}, where $D = 1, \ldots, N_{\text{p}}$.

A brief description of the proposed prediction algorithm is as follows. First, the historical traffic flow data is categorized by the day of the week, so that a nominal flow profile $u(k)$, for $k = 1, \ldots, N$ is obtained for each day of the week. Using the representative nominal flow profile $u(k)$ as the deterministic input, and the recent traffic flow $y(k)$ as the noise-contaminated output, the parameters of an ARMAX model are recursively estimated at each time step k. Based on the estimated ARMAX parameters, the Bezout equation is solved at every time step, which produces the parameters of the optimal D-step ahead output predictor $y^{\text{p}}(k+D|k)$, given only output data up to time step L and the sequence of historical data $u(k)$. In the following, we present each step in detail.

Nominal Flow Profile for each Day of the Week

In the first step, nominal historical flow profiles are determined for each day of the week. The nominal profile is robust with respect to outliers, which are caused by, for example, traffic accidents or detector failures.

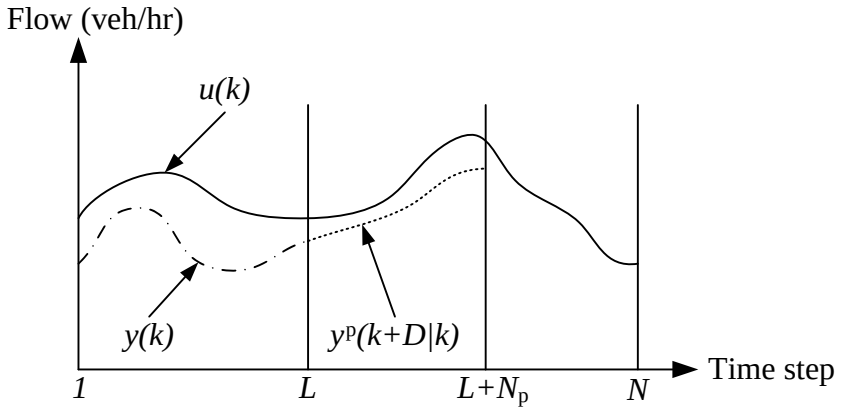


Figure 3.3: Concept of traffic flow prediction at time step L; dashed line: predicted traffic flow with horizon N_{p}; dash-dotted line: recent traffic flow measurements; solid line: nominal historical profile

Let a VDS measure the flow over the course of S full days. These data are represented as a set

$$M = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_S\} \tag{3.1}$$

of historical traffic flow data, where each element

$$\mathbf{x}_i = \big[x_i(1), x_i(2), \cdots, x_i(N)\big]^{\mathrm{T}} \in \mathbb{R}_+^N \tag{3.2}$$

is a traffic flow profile over day i, for $i = 1, \ldots, S$, with $x_i(k)$ as the flow measured at time step k.

The data are then categorized by the day of the week. Define the nominal flow profile of a given day of week as $u_d(k)$, where $d \in \{\text{Mon, Tue, Wed, Thu, Fri, Sat, Sun}\}$. The nominal flow profile of each day of the week is calculated as the median of each sampling time step. For example, the typical profile of Monday is calculated as

$$u_{\mathrm{Mon}}(k) = \text{median}\{x_i(k) : \text{day } i \text{ is a Monday}\} \ . \tag{3.3}$$

Aggregating many profiles to few typical profiles filters out some of the noise of the data, which leads to a robust estimation of typical profiles. Furthermore, by using the median, the effect of outliers is eliminated.

In this aggregation step, potentially a large amount of data are aggregated, which might require significant computational resources. However, this is not an issue for the on-line operation of the system, because the typical profile can be computed off-line.

For the on-line forecast procedure that will be described in the next section, the nominal profile of the respective day of the week is used, i.e. if the forecast day is a Monday, then $u(k) = u_{\mathrm{Mon}}(k)$ will be used.

ARMAX Model Definition and Parameter Estimation

In order to model the correlations of flow measurements in time, an autoregressive moving average model with exogenous inputs (ARMAX) model is employed in this research. Consider a recursive ARMAX model in every time step k described by the following linear stochastic difference equation

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})w(k) \ , \tag{3.4}$$

where $y(k)$ is the measured flow of the VDS at time step k, $u(k)$ is the nominal historical flow data at time step k, and $w(k)$ is assumed to be a zero-mean innovation sequence, i.e. $E\{w(k)\} = 0$ and $E\{w(k)w(k-j)\} = 0$ for $0 < j \leq k$. $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ are scalar polynomials in the backward shift operator q^{-1} $[q^{-1}y(k) \triangleq y(k-1)]$ of orders n_a, n_b and n_c, respectively, defined by

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + ... + a_{n_a} q^{-n_a} \\ B(q^{-1}) &= b_0 + b_1 q^{-1} + ... + b_{n_b} q^{-n_b} \\ C(q^{-1}) &= 1 + c_1 q^{-1} + ... + c_{n_c} q^{-n_c} \ . \end{aligned} \tag{3.5}$$

The order of the polynomials n_a, n_b and n_c are design parameters.

As described above, the exogenous known input $u(k)$ in equation (3.4) is calculated as the median of a large number of historical flows for a unique day of the week (e.g. all Monday flows in a given month). Therefore it can be considered as a deterministic input in this model. Day to day variability in the flow $y(k)$ from the median $u(k)$ can be due to both deterministic factors (e.g. the flow on a particular day could be consistently 5 % higher than the median) or stochastic effects, such as random colored noise. A fairly general model that includes many of these effects is a state space model with input and measurement white noises

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}u(k) + \mathbf{B}_w m(k) \\ y(k) &= \mathbf{C}\mathbf{x}(k) + \mathrm{D}u(k) + v(k) \end{aligned} \tag{3.6}$$

where the input noise $m(k)$ and the measurement noise $v(k)$ are both zero mean white noises, $\mathbf{x}(k) \in \mathbb{R}^n$ is the state and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{B}_w$, and $\mathbf{C}$ are matrices of the appropriate dimensions, and D is a scalar. A well-known result in stochastic systems [32] is that, under fairly mild assumptions and stationarity, the system in equation (3.6) can also be described by the ARMAX model in equation (3.4), where in this case the zero mean innovation signal $w(k)$ is equal to the Kalman filter residual for the system in equation (3.6) and the roots of the polynomial $\bar{C}(z) = z^{n_c}C(z^{-1})$ are the closed loop poles of the Kalman filter. Similar results exist for time-varying matrices $\mathbf{A}(k)$, $\mathbf{B}(k)$, $\mathbf{B}_w(k)$, $\mathbf{C}(k)$ and gain D(k) and/or non-stationarity conditions. However in these cases, the coefficients of the polynomials $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ in Equation (3.5) are time-varying.

We now assume that the flow $y(k)$ can be adequately described in terms of the ARMAX model (3.4), using the known exogenous input $u(k)$, but will not assume that the (possibly

time-varying) coefficient of the polynomials $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ in equation (3.5) are known, nor the innovations signal $w(k)$ is measurable. The coefficients of the polynomials in (3.5) are estimated using a Recursive Least Squares (RLS) Parameter Adaptation Algorithm (PAA) [32] with forgetting factor and covariance resetting [35]:

$$\begin{aligned}\hat{A}(q^{-1}) &= 1 + \hat{a}_1(k)q^{-1} + ... + \hat{a}_{n_a}(k)q^{-n_a} \\ \hat{B}(q^{-1}) &= \hat{b}_0(k) + \hat{b}_1(k)q^{-1} + ... + \hat{b}_{n_b}(k)q^{-n_b} \\ \hat{C}(q^{-1}) &= 1 + \hat{c}_1(k)q^{-1} + ... + \hat{c}_{n_c}(k)q^{-n_c}\end{aligned} \tag{3.7}$$

The a-posteriori estimation output $\hat{y}(k)$ is given by

$$\hat{y}(k) = -\hat{A}^*(q^{-1})y(k) + \hat{B}(q^{-1})u(k) + \hat{C}^*(q^{-1})e(k) \tag{3.8}$$

$$e(k) = y(k) - \hat{y}(k)\ , \tag{3.9}$$

with $\hat{A}^*(q^{-1}) = \hat{A}(q^{-1}) - 1$ and $\hat{C}^*(q^{-1}) = \hat{C}(q^{-1}) - 1$. The parameters are updated in order to make the residual $e(k)$ in (3.9) converge to an innovation sequence. Details regarding the implementation and convergence properties of the recursive parameter estimation algorithm used in this study can be found in many adaptive signal processing and control texts, such as [61, 32]. In essence, under stationary assumption and constant coefficients, it can be shown that the output estimation error $e(k) = y(k) - \hat{y}(k)$ converges to an innovation sequence and, under further assumptions regarding $C(q^{-1})$, the coefficient estimates converge to the true coefficients in (3.5).

The algorithm in [35] uses two design parameters, the forgetting factor and the regularization factor, to improve the parameter estimation performance and allow tracking of slowly time-varying coefficients. The forgetting factor is a real number greater than zero and less than one (typically very close to one), which enhances the ability of the PAA to track parameter variations. The smaller the forgetting factor is, the smaller the contribution of older data is to the RLS PAA. The regularization factor is another RLS PAA design parameter. As detailed in [35], the regularization procedure prevents the RLS gain matrix (also known as the covariance matrix) from becoming unbounded under lack of persistence of excitation and prevents the estimated parameter-bursting phenomenon. Estimated parameter bursting (i.e. parameters suddenly become large in magnitude) may lead to large overshoots in the optimal predictor.

Optimal Traffic Flow Predictor

In this section, we describe the optimal prediction method that is employed in this study. Assume that the RLS ARMAX parameter estimates in (3.7) converge such that $e(k) = y(k) - \hat{y}(k)$ converges to an innovation sequence and, in particular, the estimated noise polynomial $\hat{C}(q^{-1})$ is asymptotically stable (i.e. all roots of the polynomial $\hat{C}(z)$ lie outside the unit circle). Given the median historical flow data $u(k)$ for $k = 1, \cdots, N$ and recent flow

data $[y(1), \cdots, y(L)]$, the optimal (minimum variance) D-step ahead predictor of $y(L+D)$, denoted as $y^{\mathrm{p}}(L+D|L)$, satisfies the following difference equation [32],

$$\hat{C}(q^{-1})y^{\mathrm{p}}(k+D|k) = \hat{G}(q^{-1})y(k) + \hat{F}(q^{-1})\hat{B}(q^{-1})u(k+D) \ . \tag{3.10}$$

The polynomials $\hat{G}(q^{-1})$ and $\hat{F}(q^{-1})$

$$\hat{F}(q^{-1}) = 1 + \hat{f}_1 q^{-1} + ... + \hat{f}_{D-1} q^{-(D-1)} \tag{3.11}$$

$$\hat{G}(q^{-1}) = \hat{g}_0 + \hat{g}_1 q^{-1} + ... + \hat{g}_{n_a-1} q^{-n_g} \tag{3.12}$$

with $n_g = \max(n_c - D, n_a - 1)$, are the unique solution of the Bezout equation,

$$\hat{C}(q^{-1}) = \hat{F}(q^{-1})\hat{A}(q^{-1}) + q^{-D}\hat{G}(q^{-1}) \ . \tag{3.13}$$

Note that $\hat{F}(q^{-1})$ and $\hat{G}(q^{-1})$ depend on k and D. We omitted the indexes for legibility.

Let $\hat{H}(q^{-1})$ be the product of the two polynomial $\hat{F}(q^{-1})$ and $\hat{B}(q^{-1})$:

$$\hat{H}(q^{-1}) = \hat{F}(q^{-1})\hat{B}(q^{-1}) = \hat{b}_0 q^{-1} + (\hat{b}_1 + \hat{f}_1\hat{b}_0)q^{-2} + \cdots + \hat{f}_{D-1}\hat{b}_{n_b} q^{-(D-1+n_b)} \ . \tag{3.14}$$

Then, the optimal predictor in (3.10) can be written in transfer function form as

$$y^{\mathrm{p}}(k+D|k) = \frac{\hat{G}(q^{-1})}{\hat{C}(q^{-1})}y(k) + \frac{\hat{H}(q^{-1})}{\hat{C}(q^{-1})}u(k+D) \ . \tag{3.15}$$

The block diagram of the optimal predictor in (3.15) is shown in Figure 3.4. The predictor is a two input one output system. Given the current data $y(k)$ and nominal historical flow data $u(k+D)$ as inputs, the system generates the D-step ahead prediction of $y(k)$, which is the output $y^{\mathrm{p}}(k+D|k)$. The polynomials in the transfer functions of the predictor are updated at each time step, allowing it to track time-varying changes in the parameters in equation (3.5) using both current flow and historical flow data. This adaptive property means that the proposed predictor is potentially able to predict traffic flow under irregular and incidental traffic conditions, as the following experiments will show.

3.3 Experimental Setup

In this section, we describe the experimental environment used to validate the proposed traffic flow prediction algorithm based on empirical data from a Californian freeway.

Data

Traffic flow data for six vehicle detector stations (VDS) along the freeway corridor I-15 North, near San Diego, California, were obtained through the Performance Measurement System (PeMS) [86]. Three of the six VDS are located on the mainline with identification

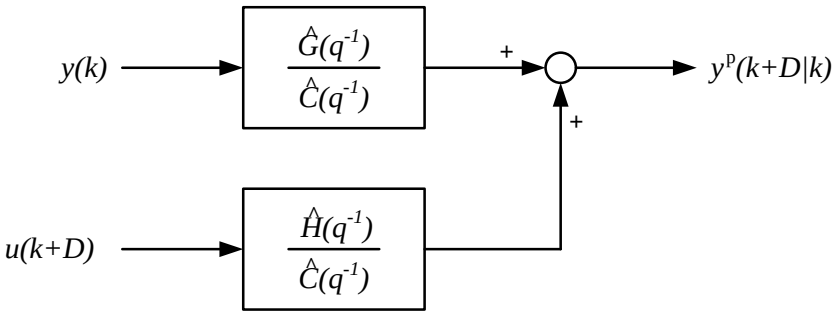


Figure 3.4: The optimal traffic flow predictor

numbers and post mileage (PM) 1108595 (at PM 24.065), 1108562 (at PM 26.249), and 1108767 (at PM 27.138). The three other VDS are located on on-ramps with identification numbers 1108596 (at PM 24.065), 1108563 (at PM 26.249), and 1108768 (at PM 27.138). Five-minute historical flow data were collected from August 1, 2012 to December 31, 2012. The collected flow data (288 observations per day) were aggregated to $\Delta t = 15$ min, which results in $N = 96$ observations per day. The five-minute historical flow data is aggregated into 15-min data to reduce the historical flow data points per day and for filtering the noise in five-minute historical data. Therefore, 15-min data were used to compute the nominal typical flow profiles $u_d(k)$ per day of the week. In order to study the influence of a special event on the traffic flow prediction method proposed in this research, we used traffic data collected during the week when the Super Bowl XLVII game took place, which is a large event in the United States. The unusual traffic flow pattern due to the Super Bowl event is shown in Figure 3.6 (solid line, $y(k)$). On Super Bowl day, people tended to stay at home or to visit friends to watch the game which began at 15:30. As a consequence, the traffic flow pattern in the period from 15:30 to 18:30 is significantly lower than the historical flow. An increase in flow that can be observed after 20:00 by the end of the game, which is probably caused by people returning to their home after watching the game. The proposed method is used to predict the flow in a rolling-horizon fashion from Monday January 28 to Sunday February 3, 2013 between 06:00 and 22:00.

ARMAX Estimation and Prediction Parameter Values

The ARMAX parameters used for the experiment are as follows. The order of ARMAX model was selected to be $(n_a, n_b, n_c) = (2, 1, 2)$. The selection of the order of the ARMAX model was based on the following three reasons. First, increasing the order of these polynomials in ARMAX model leads to overfitting and reduces the computation efficiency. Second, we chose the order of the polynomial $A(q^{-1})$ and $C(q^{-1})$ to be the same. Third, we noticed that making the orders of $A(q^{-1})$ and $C(q^{-1})$ larger than two, resulted in little or no prediction improvement. The forgetting factor was selected as 0.97, and the regularization factor was selected as 0.01 after a brief trial an error period. Both of these factors are design parameters for the RLS PAA algorithm in [35]. Since the sampling rate Δt is 15 min and the prediction

horizon is one hour ($N_\mathrm{p} = 4$), the prediction step D is varied from 1 to 4, in order to make 15 min, 30 min, 45 min and 60 min ahead predictions.

Prediction Accuracy Performance

Given the real flow data $y(k)$ and the predicted flow data $y^\mathrm{p}(k|k-D)$ at time step k and prediction step D, the performance was evaluated based on mean absolute percentage error (MAPE) of prediction step D for a VDS as follows

$$\mathrm{MAPE}(D) = \frac{1}{n}\sum_{k=k_0}^{k_1}\frac{|y(k) - y^\mathrm{p}(k|k-D)|}{y(k)}\,. \tag{3.16}$$

The starting and end time steps are $k_0 = \frac{06:00}{\Delta t}$ and $k_1 = \frac{22:00}{\Delta t}$, respectively, which leads to $n = k_1 - k_0 + 1 = 65$ time steps for each day where a prediction is started. This error (3.16) is further averaged over the detector type (mainline/on-ramp).

3.4 Results and Discussion

In this section, the traffic flow prediction results of the method proposed are presented and compared against a simple predictor based only on the nominal historical profile. We first show the prediction of a regular day, and then of the Super Bowl day, where irregular traffic flows occurred.

Flow Prediction on a Regular Day

The flow prediction result of a mainline VDS (1108595, 5 lanes) of a weekday (Thursday 31 January 2013) is shown in the top part of Figure 3.5. The flow measurements $y(k)$ are indicated by the solid line. The nominal historical profile $u(k)$ is shown by the dashed line. The prediction, indicated by the remaining four lines, is close to the actual measurements so that only very small prediction error occurs. Because the regular is close to the nominal historical profile, the prediction is very close to the actual data.

The flow prediction result of an on-ramp VDS (1108563) on a weekday (Wednesday 30 January 2013) is shown in the bottom part of Figure 3.5. The difference between the predicted and the actual flows is small.

Flow Prediction on Super Bowl Sunday

The flow prediction of a mainline VDS (1108595) for the Super Bowl day (Sunday 3 February 2013) is shown in the top part of Figure 3.6. Notice that flow leading up to the beginning of the game at 15:30 is drastically lower than the nominal flow $u(k)$. Also, the top part of

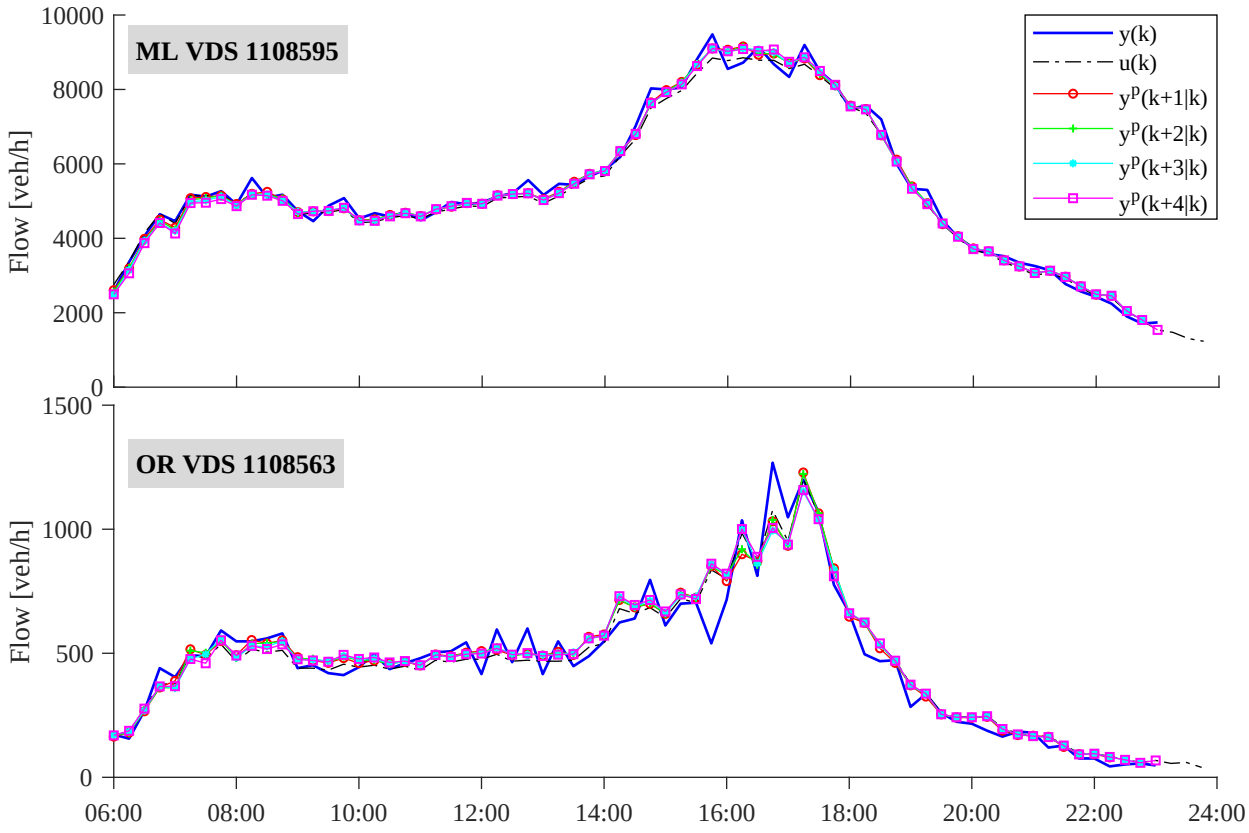


Figure 3.5: Flow prediction in weekday; top: mainline; bottom: on-ramp

Figure 3.6 shows that the prediction result $y^{\mathrm{p}}(k+D|k)$ for $D = 1, \ldots, 4$ more closely follows the measured flow $y(k)$ than the historical flow $u(k)$.

Using only the nominal historical profile as forecast leads to very large errors, since the drop of the traffic flow at 15:30 is completely ignored. The proposed prediction algorithm therefore outperforms the historical predictor. Moreover, the result shows that the one step prediction $y^{\mathrm{p}}(k+1|k)$ has better performance than larger steps. As the prediction step increases, the prediction deviates from the measurement flow, especially in the case of special event. The prediction error increases when the prediction step D increases.

The traffic flow prediction of an on-ramp VDS (1108563) on the Super Bowl day is shown in the bottom part of Figure 3.6. The influence of the Super Bowl game is visible at this location as well. The measured traffic flow $y(k)$ behaves normally until 15:30, then the flow decreases between 15:30 and 18:30. Before 15:30, the proposed prediction result $y^{\mathrm{p}}(k+D|k)$ for $D = 1, \ldots, 4$ and the historical flow $u(k)$ matches the measured traffic flow data $y(k)$. However, the prediction starts to deviate from the measured flow at 15:30. After 15:30, the prediction overestimates the traffic flow until 18:30, because the measured flow data has a significant difference with historical flow data.

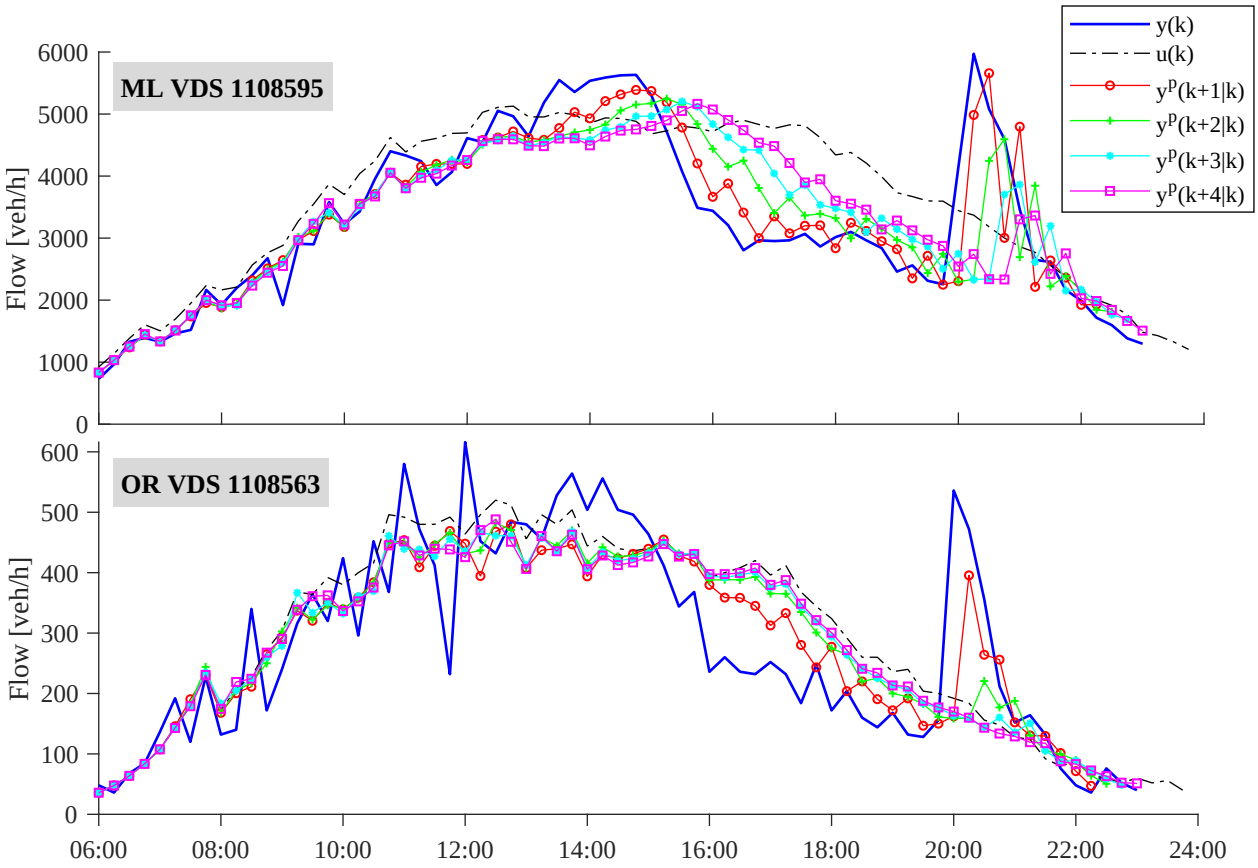


Figure 3.6: Flow prediction of Super Bowl Sunday; top: mainline; bottom: on-ramp

Evaluation of Prediction Performance Using the MAPE Criterion

The MAPE criterion defined in (3.16) are averaged for three mainline and three on-ramp VDS, which is shown in Table 3.1 and Table 3.2, respectively. In addition, the MAPE criterion evaluates the prediction performance for the entire day rather than a specific time period of a day because the predictor is using the historical pattern of a whole day to make prediction.

In the mainline case, both of the MAPE values for the proposed method and the historical predictor are approximately 7% for the normal days (Monday to Saturday), which indicates that the MAPE error does not change significantly across normal days. However, due to the influence of the special event on Sunday, the MAPE values are higher than other days in the week. On Sunday (Super Bowl day), the error of the proposed method is 10.3% for the $D = 1$ step prediction. The error increases with the length of the prediction horizon. At $D = 4$, the prediction error is 16.6%, while the MAPE error using historical data for prediction is 21.5%. The proposed method based on the ARMAX model therefore outperforms using only historical data to predict the traffic flow.

	Mainline MAPE [%]						
Predictor	Mon	Tue	Wed	Thu	Fri	Sat	Sun*
$y^{\mathrm{p}}(k+1\|k)$	6.4	5.9	6.2	5.1	6.7	6.3	10.3
$y^{\mathrm{p}}(k+2\|k)$	7.4	5.9	6.4	5.3	7.0	7.1	12.4
$y^{\mathrm{p}}(k+3\|k)$	7.8	5.9	6.7	5.5	7.0	7.2	14.5
$y^{\mathrm{p}}(k+4\|k)$	7.8	6.1	7.0	5.5	7.0	7.6	16.6
$u(k)$	8.0	7.0	7.6	5.9	7.4	7.9	21.5

*Super Bowl day

Table 3.1: Comparison of proposed predictor and historical predictor for mainline VDS

	On-ramp MAPE [%]						
Predictor	Mon	Tue	Wed	Thu	Fri	Sat	Sun*
$y^{\mathrm{p}}(k+1\|k)$	22.1	21.9	23.1	21.9	28.1	31.7	31.3
$y^{\mathrm{p}}(k+2\|k)$	21.9	21.6	23.1	21.4	27.9	32.1	32.9
$y^{\mathrm{p}}(k+3\|k)$	21.8	22.3	22.7	21.6	27.0	31.6	33.6
$y^{\mathrm{p}}(k+4\|k)$	20.6	20.6	22.8	21.1	26.4	31.8	34.3
$u(k)$	22.8	21.9	24.0	21.9	28.8	32.3	38.7

*Super Bowl day

Table 3.2: Comparison of proposed predictor and historical predictor for on-ramp VDS

In the on-ramp case, the MAPE value for both the proposed method and the historical predictor are distributed from 20% to 29% for the normal days (Monday to Saturday), which is higher than normal days in the mainline case. On Sunday (Super Bowl day), the error of the proposed method is 31.3% for the $D = 1$ step prediction. At $D = 4$, the prediction error is 34.3%, while the MAPE error using historical data for prediction is 38.7%. The proposed method based on the ARMAX model still outperforms using only historical data to predict the traffic flow in the on-ramp case.

The predictions at mainline show a systematically lower error than those at on-ramps. The MAPE values in Table 3.1 and Table 3.2 show an error for mainline flow prediction between 6% and 8%, while the on-ramp flow prediction error ranges from 20% to 33%, for a regular day. The reason is that mainline traffic is significantly higher than that of on-ramps and is, in fact, the aggregation of many single on-ramps flows, namely those that are located upstream. This aggregation leads to a systematic reduction of signal noise. Since the signal is less corrupted by noise, the estimation of the ARMAX parameters is more precise and consequently the prediction is more precise, i.e. its errors are lower.

3.5 Summary

A simple and fast, yet accurate, method to forecast traffic flow based on data from vehicle detector stations have developed in this research. An optimal multi-step ahead predictor combines the most recent measured flow data with a nominal historical flow profile. The parameters of this predictor are estimated by an autoregressive moving average model with exogenous input (ARMAX). This method predicts traffic flow under irregular conditions significantly better than a simple predictor based on the nominal historical profile alone. We showed by the example of the Super Bowl day, where traffic patterns differ drastically from regular Sundays, that the prediction error is reduced by up to ten percentage points.

The nominal historical profile is computed as the median over months of historical data, stratified by the day of the week. Using the median ensures that outliers caused by failing detectors or irregular days do not affect the nominal profile. The nominal profiles are computed off-line so that the on-line forecast only consists of the ARMAX estimation and prediction method, which is computationally very fast. This combination of a robust estimation of the nominal profile, the accurate prediction of the near-future traffic flow, and the fast computation make this method suitable for practical applications of boundary flow forecasting and its use in traffic management systems.

Chapter 4

Microscopic Traffic Flow Model Calibration

4.1 Introduction

The research in this chapter demonstrates the first stage of a larger project that has as its goal to design and implement a coordinated ramp metering (CRM) traffic responsive control system for the California State Route 99 Northbound (SR-99N) in Sacramento, California. The simulation model of SR-99N will be used as a test bed for evaluating the performance of different control strategies [66] before the field implementation presented in Chapter 5.

The purpose of building traffic simulation models is to represent the existing traffic condition and predict future conditions when new traffic management alternatives are introduced. Model calibration is an iterative procedure by which the user adjusts the parameter of the model so that the simulation results can match the field traffic conditions. Based on the level of modeling detail, traffic simulation models can be generally classified into microscopic and macroscopic model. Macroscopic models describe the traffic stream characteristics, such as flow, speed, and density. Microscopic models describe the individual vehicle-driver behavior and their reactions to other drivers and environments. Compared to a microscopic model, a macroscopic model is normally easier to calibrate because its model parameters flow, speed, and density can be directly obtained from existing vehicle detector infrastructure. The main difficulty of calibrating a microscopic model is that its model parameters contain not only traffic stream characteristic, but also driver behavior parameters and more detailed information about the system. A unified methodology for calibrating a microscopic traffic flow model had not been proposed, prior to the results presented in this Chapter.

Several systematic procedures [28] and guidelines [39] [26] have been proposed to improve the performance of the model calibration process for reproducing the traffic condition by simulation. Some researchers calibrate the model based on original-destination (O-D) demand estimation. Some researchers calibrate the model based on adjusting driver behavior parameters such that the simulation result and field measured flow, speed, or occupancy data can

reasonably agree with each other [31] [65]. One difficulty of model calibration is that because the microscopic model contains many coupled car-following models as its submodels and the relation between model parameters of car-following models and the aggregate output of the microscopic model are very complex, it is impossible to obtain a closed-form mathematical representation of the simulation model in terms of model parameters. Mathematical optimization methods, such as gradient search methods, cannot be applied on this case due to the lack of a closed-form model. In order to overcome this difficulty of model calibration, researchers have proposed simulation-based optimization method for model calibration [36], [9], [19], [67], which is a model calibration technique by searching the optimal combination of model parameters so that the objective function (usually, it is a function that represents the discrepancy between simulation output and field measurement) is minimized.

Another difficulty of model calibration is computational time, especially for microsimulation models of large networks. The microsimulation model contains a large number of adjustable parameters and the model input contains uncertainty, which is due to the error of field measurement and the stochastic nature of traffic flow. In order to reduce the number of iterations of the model calibration process, a small group of parameters and model inputs which have a significant influence on the model inputs can be identified by sensitivity analysis [88]. The unimportant parameters of the model can be left out of the optimization process and therefore the iteration time of model calibration can be reduced. The work in [93] studied the global sensitivity analysis of the Newell and Gipps' car following-model and it indicates that the fuel consumption and pollutant emissions output of the microsimulation model are sensitive to the parameters of the car-following model.

This chapter presents a procedure for constructing and calibrating a microsimulation model of a congested freeway with multiple vehicle classes by using AIMSUN, a popularly-used commercial microsimulation software [2]. This procedure is applied to a 13 mile long stretch of SR-99 Northbound in Sacramento, California. The test site contains a mainline high occupancy vehicle (HOV) lane, 16 metered on-ramps with and without HOV lanes, freeway interconnection and three interacting bottlenecks. The model construction and calibration procedure is as follows: (1) building the road geometry in AIMSUN, (2) collection of the traffic condition data from the PeMS database, (3) imputation of missing data, (4) estimation of on-ramp demand and off-ramp turning ratio, (5) identification of recurring bottlenecks, (6) setting model parameters, and (7) adjustment of model parameters such that the simulation results match with the field measurements. Model calibration results indicate that the simulated flow match with the field data near bottleneck locations reasonably well under the criterion of root mean square percentage error (RMSPE) and GEH criterion. The calibrated AIMSUN microsimulation model was subsequently used as a simulation test bed for implementing and testing a coordinated ramp metering (CRM) control strategy, which has the goal of alleviating the heavy congestion of the SR-99N freeway during the morning peak period. These results are discussed in Chapter 5.

4.2 A Case study: SR-99 Northbound

Network Configuration

The site of the simulation study is the Northbound direction of SR-99 (State Route 99) from Elk Grove Blvd at absolute postmile 287.23 to the freeway interchange point of SR-99N and US-50 (U.S. Route 50) at absolute postmile 298.5. The network configuration is shown in Figure 4.1, where the numbered black boxes, red circles and green circles label the indices of mainline VDS, off-ramp VDS and on-ramp VDS, respectively. The studied time period is between 5:00 am and 11:00 am. The test site is an 11-mile stretch of freeway that sustains heavy congestion during the morning commute. Congestion usually begins around 6:30 am, peaked at 8:00 am, and finally dissipated at around 10:00 am. The HOV lane in the leftmost lane spans the entire mainline of the test site. The HOV lane is reserved for two or more passengers per vehicle and its operational time is from 6:00 am to 10:00 am and from 3:00 pm to 7:00 pm. The test site has 16 on-ramps and 11 off-ramps. Each of these 16 on-ramps is equipped with ramp metering, which has a corresponding upstream mainline and downstream on-ramp detector station for traffic responsive control. Eleven of these 16 on-ramps have HOV bypass lanes, which are not metered by ramp metering.

The construction of a realistic simulation model requires a complete and detailed description of the road geometry of the site. The information needed to build the model contains freeway curvature, the number of lanes, the location of lane drops, bottleneck locations, weaving sections, the location of on-ramps and off-ramps, and the length of on-ramps and off-ramps. In addition, the measurement and control facility of the freeway also needs to be built in the model. They are the locations of vehicle detector stations, the locations of queue detectors and the location of ramp metering signals. The geometric information used for this study is obtained from `OpenStreetMap` [76] and Google Maps [33]. The facility information about the test site is provided by PeMS [86] and the Caltrans District 3 Ramp Metering Group.

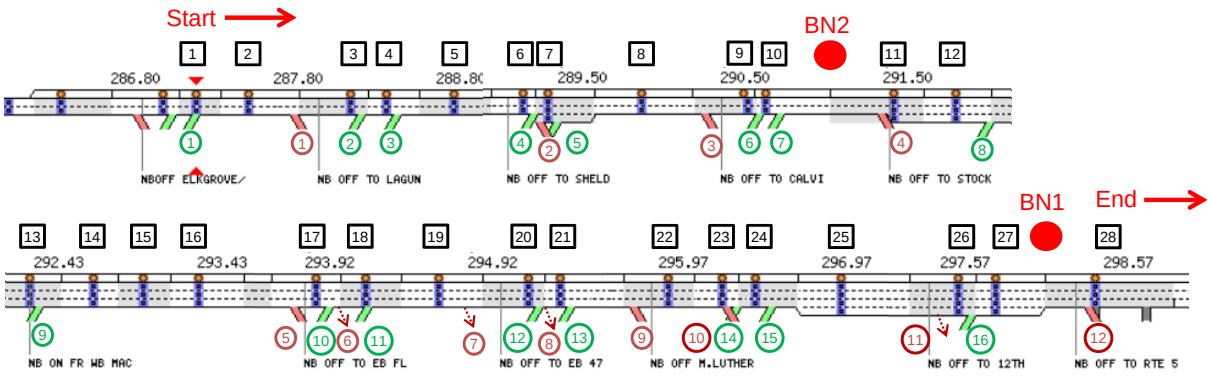


Figure 4.1: SR-99N network configuration and VDS deployment.

Index	Postmile	Name	HOV	GP
ML 1	287.295	Elk Grove Blvd	315828	313190
ML 2	287.615	Red Fox	315831	314697
ML 3	288.25	EB Laguna Blvd	315827	313166
ML 4	288.47	WB Laguna Blvd	315823	313172
ML 5	288.88	Laguna Creek	315826	313178
ML 6	289.274	EB Sheldon Rd	317956	317960
ML 7	289.423	WB Sheldon Rd	317947	317948
ML 8	290.003	Jacinto Rd.	315834	314625
ML 9	290.65	EB Calvine Rd.	315841	312648
ML 10	290.763	WB Calvine Rd.	315842	312651
ML 11	291.549	Stockton Blvd	315836	312233
ML 12	291.93	EB Mack Rd	315838	312382
ML 13	292.38	WB Mack Rd	315822	312386
ML 14	292.77	NameTangerine Ave.	315843	312388
ML 15	293.08	Pomegranate Ave	317909	317910
ML 16	293.42	Orange Ave	315847	312421
ML 17	293.97	EB Florin Rd	315849	312422
ML 18	294.27	WB Florin Rd	315850	312425
ML 19	294.717	Turnbridge Dr	315852	312513
ML 20	295.27	EB 47th Ave	315853	312514
ML 21	295.47	EB 47th Ave	315873	312520
ML 22	296.007	Martin L. King Jr	315854	312523
ML 23	296.335	EB Fruitridge	315855	312525
ML 24	296.54	WB Fruitridge	315856	312527
ML 25	297.07	21st Ave UC	317895	317896
ML 26	297.655	12th Ave	315825	312562
ML 27	297.89	8th Ave. POC	318565	318566
ML 28	298.5	Broadway	318576	318575

Table 4.1: List of mainline VDS

Index	Postmile	VDS	Name	Number of lanes	Number of HOV lanes
OR 1	287.23	314107	Elk Grove Blvd	3	1
OR 2	288.25	314114	EB Laguna Blvd	1	0
OR 3	288.47	314098	WB Laguna Blvd	1	0
OR 4	289.274	317959	EB Sheldon Rd	2	1
OR 5	289.423	317949	WB Sheldon Rd	3	1
OR 6	290.662	312649	EB Calvine Rd.	2	1
OR 7	290.791	312652	WB Calvine Rd.	3	1
OR 8	292.084	312383	EB Mack Rd	2	1
OR 9	292.383	312387	WB Mack Rd	2	1
OR 10	293.986	312423	EB Florin Rd	1	0
OR 11	294.217	312426	WB Florin Rd	2	1
OR 12	295.27	312515	EB 47th Ave	2	1
OR 13	295.47	312521	WB 47th Ave	2	1
OR 14	296.366	312526	EB Fruitridge	1	0
OR 15	296.581	312528	WB Fruitridge	1	0
OR 16	297.679	312563	12th Ave	1	0

Table 4.2: List of on-ramp VDS

Index	Postmile	VDS	Name	Number of lanes
FR 1	287.974	314115	EB Lanuna Blvd	2
FR 2	288.96	317961	Eb Sheldon Rd	3
FR 3	290.454	312650	EB Calvine Rd	1
FR 4	291.539	314615	Stockton Blvd	1
FR 5	293.896	312424	EB Florin Rd	1
FR 6	294.07	-	WB Florin Rd	1
FR 7	295.06	-	EB 47th Ave	1
FR 8	295.34	-	WB 47th Ave	1
FR 9	295.861	312524	Martin L. King Jr	1
FR 10	296.426	312529	WB Fruitridge	1
FR 11	297.44	-	12th Ave	1
FR 12	298.5	318577	99NB to I-50	2

Table 4.3: List of off-ramp VDS

Traffic Data Sources

The data source used in this traffic simulation and model calibration is obtained from the PeMS data base. The traffic flow and occupancy data are collected from the PeMS data base in 5 minute intervals for mainline, on-ramp and off-ramp detectors at the target site. This network contains 28 general purpose lane VDS, 28 HOV lane VDS (Table 4.1), 16 on-ramp VDS (Table 4.2), 11 off-ramp VDS and one freeway to freeway interconnection VDS (Table 4.3). That freeway to freeway connection VDS is modeled as an off-ramp in the model. Four VDS are missing among 11 off-ramps. One on-ramp VDS has no data. For the mainline data, the flow conservation law is a criterion to check the data credibility. Seven out of 28 general purpose lane detectors (25%) are problematic because the flow data from those detectors indicate significant incompatibility with upstream and downstream measurements. Six out of the 28 HOV lane detectors (22%) are problematic. Those malfunctioning detectors are found by comparing flow measurements at consecutive detectors upstream and downstream. For the missing on-ramp and off-ramp data, if their immediately upstream or downstream mainline detector is available, then the flow conservation law is used for data imputation. The missing on-ramp flow data r_i and missing off-ramp flow data s_i can be calculated by

$$r_i = f^H_{DN,i} + f^G_{DN,i} - f^H_{UP,i} - f^G_{UP,i} \tag{4.1}$$

$$s_i = f^H_{UP,i} + f^G_{UP,i} - f^H_{DN,i} - f^G_{DN,i} \tag{4.2}$$

, where r_i is the imputed flow data of on-ramp detector i, s_i is the imputed flow data of detector i, $f^H_{DN,i}$ and $f^G_{DN,i}$ are the flow measurement of HOV lane and flow measurement of general purpose (GP) lanes at the immediate downstream of detector i respectively, $f^H_{UP,i}$ and $f^G_{UP,i}$ are the flow measurements of HOV lane and flow measurement of GP lane at the immediate upstream of detector i respectively.

Traffic Demand Data

Traffic demand data is the input for the microsimulation model. The demand data in AIMSUN can be either O-D matrix or traffic states. This research uses traffic states as the form of demand and therefore the user needs to specify the input flow (boundary flow) for all vehicle classes at each entrance section (the most upstream mainline and all on-ramps) and the turning ratio at each exit section (the most downstream mainline and all off-ramps). Three types of vehicles, general purpose vehicle (GPV), high occupancy vehicle (HOV) and truck, are added into this model. The demand data consist of two parts: demand at the most upstream mainline and demand at each on-ramp. The most upstream mainline demand is composed of single occupancy vehicle flow, high occupancy vehicle demand and truck demand, which can be represented as

$$d_0 = d_G + d_H + d_T \tag{4.3}$$

where d_0 is the total demand at the most upstream mainline location in the model, d_G, d_H, d_T are the demand of general purpose vehicle, high occupancy vehicle and truck at the most

upstream mainline location, respectively. In this research, the percentage of truck demand P_T is assumed to be 6.3% and the Passenger Car Equivalent (PCE) value c of a truck is assumed to be 3, therefore, the demand of each type of vehicle is given by

$$d_G = (1 - P_T c) f_{G,0} \tag{4.4}$$
$$d_H = f_{H,0} \tag{4.5}$$
$$d_T = P_T c f_{G,0} \tag{4.6}$$

where $f_{G,0}$ and $f_{H,0}$ are the flow measurements of general purpose lane and HOV lane at the most upstream end of the target site respectively.

In order to model the high occupancy vehicles and the deployment of ramp metering in the field, a single on-ramp is modeled as the set of lanes, which contains a high occupancy vehicle lane and single vehicle lanes. Therefore, the on-ramp demand d_i for on-ramp i is composed by

$$d_i = d_{H,i} + d_{G,i} \tag{4.7}$$

where $d_{H,i}$ and $d_{G,i}$ are the on-ramp demand of high occupancy vehicle and of general purpose vehicle lanes at on-ramp i, respectively. The on-ramp demand is assumed to be α percent increasing of the on-ramp flow measurement, that is

$$d_{H,i} = (1 + \alpha) f_{H,i} \tag{4.8}$$
$$d_{G,i} = (1 + \alpha) f_{G,i} \tag{4.9}$$

where f_H, i and f_G, i are the flow measurements in the HOV lane and GP lane respectively. $\alpha = 5\%$ in this research. (Remark: The data source from PeMS is the detected flow instead of demand. In free-flow or no-queue situation, one could use the flow as the demand; but once there is a queue, or even worse, the onramp queue spills back to the arterial/surface street, the detected flow is not the demand anymore. Therefore, the simulation increased the demand 5% in order to simulate the recurrent congestion pattern.)

The output flow of the model consists of two parts: output flow at the most downstream boundary and output flow at each off-ramp. This research uses turning ratio at each exit of the model to quantify the output flow. The user needs to define the turning ratio for each type of vehicle in the model. However, there are no vehicle detector stations specific to HOVs at the off ramps and thus no measurement for the HOV exiting flow. Total off-ramp flow data is the only measurement at each off-ramp. In order to model the off-ramp HOV flow, a certain proportion P of the total off-ramp flow is assumed to be HOV flow. That is

$$s_i = s_i^H + s_i^G \tag{4.10}$$

where $s_i^H = P s_i$ is the HOV flow at off-ramp i, $s_i^G = (1-P)s_i$ is the GP flow at off-ramp i, and p is the percentage of HOV at off-ramp i and it is assumed to be 20% in this research, which is determined by experience and model calibration result. Besides, the flow measurements

immediately upstream and downstream of an off-ramp are needed for calculating the turning ratio. Since the mainline detectors are usually close to the off-ramps (about 0.2 mile), it is reasonable to assume that exiting HOVs on the mainline have changed to the general purpose lane. Then, the total HOV flow $f^H_{total,i}$ from the immediately upstream mainline of the off-ramp i contains HOVs in the HOV lane and HOVs in the general purpose lane, that is

$$f^H_{total,i} = f^H_{UP,i} + s^H_i = f^H_{UP,i} + Ps_i \tag{4.11}$$

where $f^H_{UP,i}$ is the flow measurement in the HOV lane immediately upstream of off-ramp i. By flow conservation law, the total GP flow $f^G_{total,i}$ from the immediately upstream mainline of the off-ramp i becomes

$$f^G_{total,i} = f^G_{UP,i} - s^H_i = f^G_{UP,i} - Ps_i \tag{4.12}$$

where $f^G_{UP,i}$ is the flow measurement of the upstream detector in the GP lane immediately upstream of off-ramp i. Therefore, the turning ratio of HOV β^H_i at off-ramp i can be calculated by

$$\beta^H_i = \frac{s^H_i}{f^H_{total,i}} = \frac{Ps_i}{f^H_{UP,i} + Ps_i} \tag{4.13}$$

and the turning ratio of GPV β^G_i at off-ramp i can be calculated by

$$\beta^G_i = \frac{s^G_i}{f^G_{total,i}} = \frac{(1-P)s_i}{f^G_{UP,i} - Ps_i}. \tag{4.14}$$

Identification of Recurrent Bottlenecks

It is necessary to study the causes of congestion on SR-99N and to identify the recurrent bottleneck locations before the model calibration process. This analysis provides guidance in tuning model parameters. By studying the speed contours from February, March, September and October in 2013 and 2014, a representative congestion pattern is selected and it is shown in Figure 4.2. Three bottleneck locations are identified, which are:
C(1): Near the upstream of SR-99N and US-50 interchange point (PM 297.65)
C(2): Near Fruitridge Road (PM 296.54), and
C(3): Near Calvine Road (PM 290.76)

The causes of congestion near each bottleneck are proposed to be the following: C(1) at PM 297.65: downstream congestion caused by diverging traffic to US 50 EB and WB. This congestion may propagate back to upstream bottleneck at PM 296.54. C(2) at PM 296.54: middle congestion caused by merging traffic from two on-ramps from EB and WB Fruitridge Rd. These two on-ramps are only 350 meters apart. The congestion is light near this location, but it becomes more severe when the on-ramp demand is high. The congestion

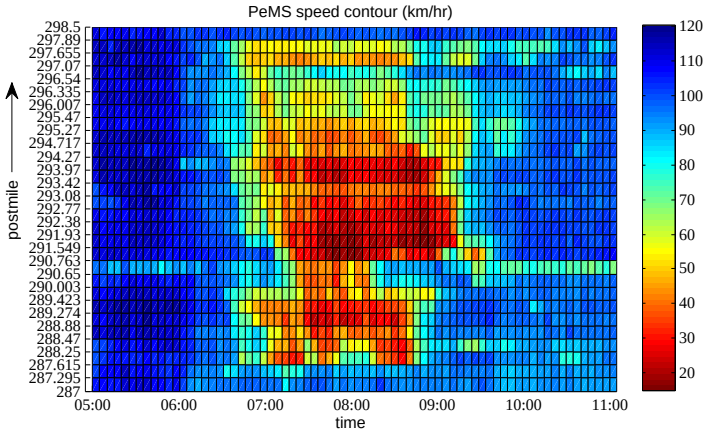


Figure 4.2: Morning speed contour on March 19, 2015, arrow indicates direction of traffic.

may propagate upstream and passes over the on-ramps EB and WB 47th Ave near PM 295.47. C(3) at PM 290.76: upstream congestion caused by merging traffic from EB and WB Calvine Rd.. These two on-ramps are only 320 meters apart and the demand from Calvine Rd. is very high. The EB and WB Calvine Rd. on-ramp flow could exceed 600 veh/hr (total flow of two lanes) and 1200 veh/hr (total flow of three lanes) during the morning peak, respectively. Since the distance between first two bottlenecks C(1) and C(2) located in downstream is only about two miles apart and these two bottlenecks will merge into one bottleneck if the congestion in the most downstream location propagates upstream, it is reasonable to model the first two bottlenecks C(1) and C(2) as one bottleneck in simulation. The combined downstream bottleneck is denoted as BN1 at PM 297.65, and the upstream one is denoted as BN2 at PM 290.76.

4.3 Model Calibration

The goal of model calibration is to find a set of parameters and inputs of the simulation model such that the output of the model can correctly reproduce observed data. Model calibration can be formulated as a numerical optimization problem [105], which can be written as

$$\begin{aligned} \underset{\alpha,\beta}{\text{minimize}} \quad & f(y^s; y^o) \\ \text{subject to} \quad & y^s = s(\alpha, \beta; u, d, g) \\ & \alpha^l \leq \alpha \leq \alpha^u \\ & \beta^l \leq \beta \leq \beta^u \end{aligned} \tag{4.15}$$

where y^o is observed data, which includes field flow measurement, field occupancy measurement and other traffic state measurements obtained from the field, y^s is simulation output, which includes flow, speed, occupancy, density and other traffic state information generated

by the simulation model, s is the simulation model, α is the global parameter of driver behavior, which includes reaction time, reaction time at stop, queueing up/leaving speed, percent over take, percent recover and vehicle attribute parameters of each vehicle type (single occupancy vehicle, high occupancy vehicles, truck, or other vehicle type), which includes maximum desired speed, maximum acceleration/deceleration and other parameters. β is the local parameter of driver behavior, which includes section maximum speed, lane speed limit, turning speed, look ahead distance and other parameters. α^u and β^u are upper bounds of model parameters, α^l and β^l are lower bounds of model parameters. u is control input of model, which includes variable speed limit or ramp metering signal, d is traffic demand data, which can be represented by O-D matrix or boundary flow (the most upstream flow and all on-ramp flow) and turning ratio, and g is the geographical parameter of the model, which includes length of road section, length of on-ramps, number of lanes and other parameters. The objective function f for model calibration is a measurement of goodness of fit that represents the discrepancy between simulation output and observed data (field data). One common measurement of goodness of fit is root mean square percentage error (RMSPE), which is defined by

$$\text{RMSPE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\frac{x(k)-y(k)}{y(k)}\right)^2} \tag{4.16}$$

where N is the number of field measurements, $x(k)$ is the simulation data and $y(k)$ is the field measurement at time k. Another objective function can be selected as GEH criterion, which is used to evaluate the flow calibration performance. According to FHWA guidelines [28], the GEH criterion is given by

$$\text{GEH} = \sqrt{\frac{2(x(k)-y(k))^2}{x(k)+y(k)}} \tag{4.17}$$

where $x(k)$ is the model simulated traffic volume, and $y(k)$ is the field traffic count. GEH values below 5 are considered a good match between estimated and observed counts and at least 85% of the observed links in a traffic model should have a GEH less than 5. If the GEH is greater than 10, there could be an error in the demand model or the field data.

Since the microsimulation model is constructed in AIMSUN (or other commercial traffic simulation software) and the model itself is a large-scale and complex traffic network, the model s may not necessarily be expressed by an analytical form, both of the objective function f and model s become in the form of a black-box model. The way to solve this optimization problem in this research is that global parameters are adjusted in the first stage until the flow and the traffic pattern in speed contour roughly match with the field data, then the local parameters are adjusted. The calibration parameters that have been considered in this research are shown in Table 4.4, and the description of each parameter can be referred to [1]. The selected parameters for car, HOV and truck are shown in Figure 4.3.

	Parameter name	Range	Unit
Global	Minimum mean headway	0.9∼1.2	sec.
	Reaction time for GPV/HOV	0.7∼ 1.5	sec.
	Reaction time for truck	1.5∼ 3	sec.
	Queue entry speed	1∼1.5	m/s
	Queue exit speed	4∼8	m/s
Local	Lane changing cooperation	70∼100	%
	Reaction time variation	0∼3	unitless
	Distance zone 1	0.5L∼0.7L	m
	Distance zone 2	0.5L∼0.2L	m

Note: L is the length of a section.

Table 4.4: Calibration range of parameters

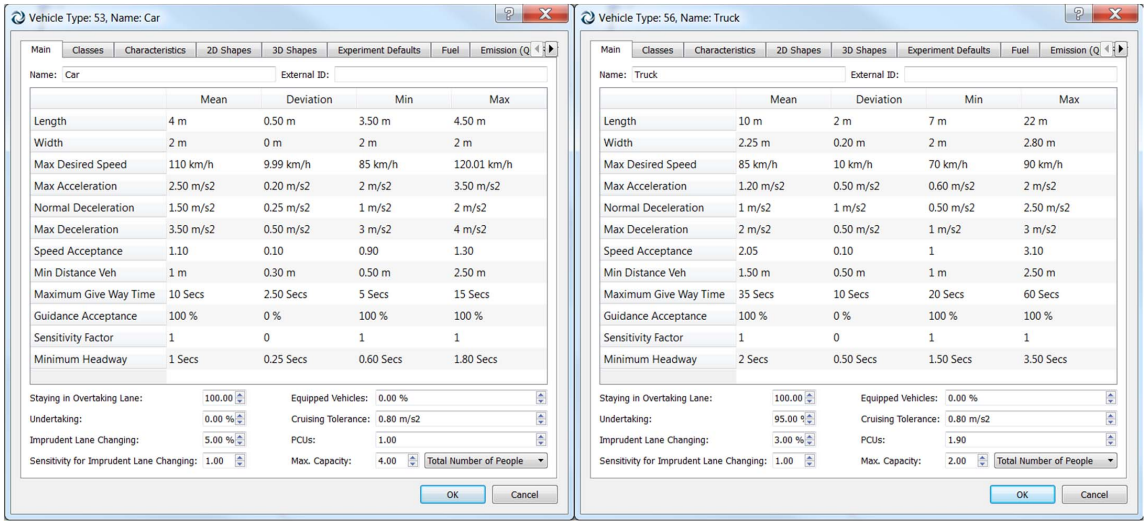


Figure 4.3: Parameter selection, left: car and HOV parameters, right: truck parameters

4.4 Results and Discussion

In order to evaluate model calibration performance, several important locations are selected for comparing the simulation results with the measured field data. 5 minutes data is used for model calibration. These selected locations can be classified into three categories: (1) mainline locations near BN1 and BN2, (2) each on-ramp and (3) each off-ramp. For mainline locations near BN1 (the downstream bottleneck), VDS group index 25, 20, 17, and 12 are selected for observing the traffic behavior near BN1. For mainline locations near BN2 (the upstream bottleneck), VDS group index 9, 8, 6, and 2 are selected. Each on-ramp and each off-ramp location are used to verify that the input and output flow are consistent with the field data, which also helps to detect problems during the model calibration process.

Traffic volume variation is different from one simulation run to the next if the random number seed is changed, which is like what happens in reality from one day to the next. This stochastic nature of traffic flow can lead to excessive variations if some of the underlying model parameters are not properly calibrated. Especially in modeling a congested freeway with multiple bottlenecks or multiple vehicle classes, improper model parameters could result in abnormal car-following behavior or lane-changing decisions and therefore cause unexpected congestion. In order to check that the model parameters are well calibrated, ten replications using different random number seeds are carried out after model calibration. The result of ten replications indicates that the traffic volume does not have excessive variation and the simulation results are the average of ten replications.

Figure 4.4 shows the mainline flow comparison between simulation and field measurements. The model calibration errors for BN1 and BN2 are listed in Table 4.5. The results indicate that the BN2 (upstream) has better model calibration performance in RMSPE and GEH value than BN1 (downstream). The error in VDS ML 12 is due to the flow error in off-ramp 4. The off-ramp flow of the model is obtained by assigning the turning ratio calculated by equation (4.13) and (4.14) for each off-ramp. The flow measurement outliers in the adjacent VDS of the off-ramp could induce a sudden change of split ratio and therefore turning ratio error induces off-ramp flow error. The error in VDS ML 17 is due to three missing off-ramps (FR 6, FR7 and FR8) in its downstream. The flow data of those missing off-ramps are recovered by equation (4.2) and all off-ramp flow data are shown in Figure 4.5. The off-ramp 7 has large error and therefore it induces the VDS ML 17 error.

Figure 4.6 shows the on-ramp flow comparison between simulation and field measurement. It indicates that all simulated on-ramp flow data match with the field data very well except for on-ramp 7. The error in on-ramp 7 may be caused by the on-ramp data outliers. The model calibration errors for on-ramps and off-ramps are listed in Table 4.6.

VDS BN1	RMSPE	GEH	VDS BN2	RMSPE	GEH
ML 25	16.26 %	75 %	ML 9	12.32 %	90.28 %
ML 20	14.19 %	86.11 %	ML 8	10.95 %	88.89 %
ML 17	14.57 %	77.78%	ML 6	12.61 %	87.5 %
ML 12	16.56 %	73.61%	ML 2	8.1 %	90.28 %

Table 4.5: Calibration error of BN1 and BN2

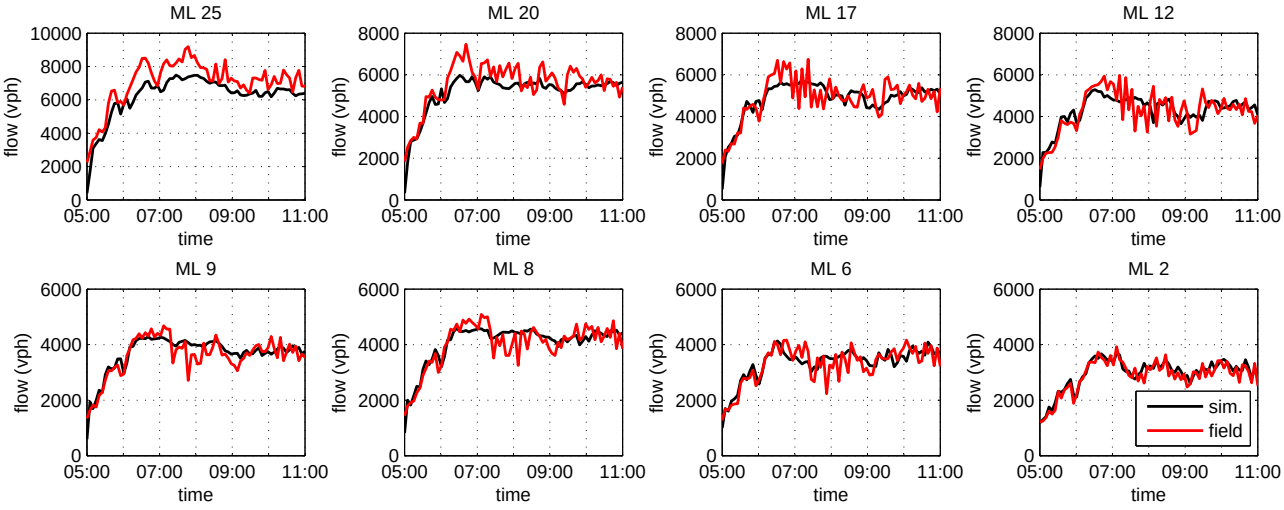


Figure 4.4: Simulated and measured flow near BN1 (top) and BN2 (bottom)

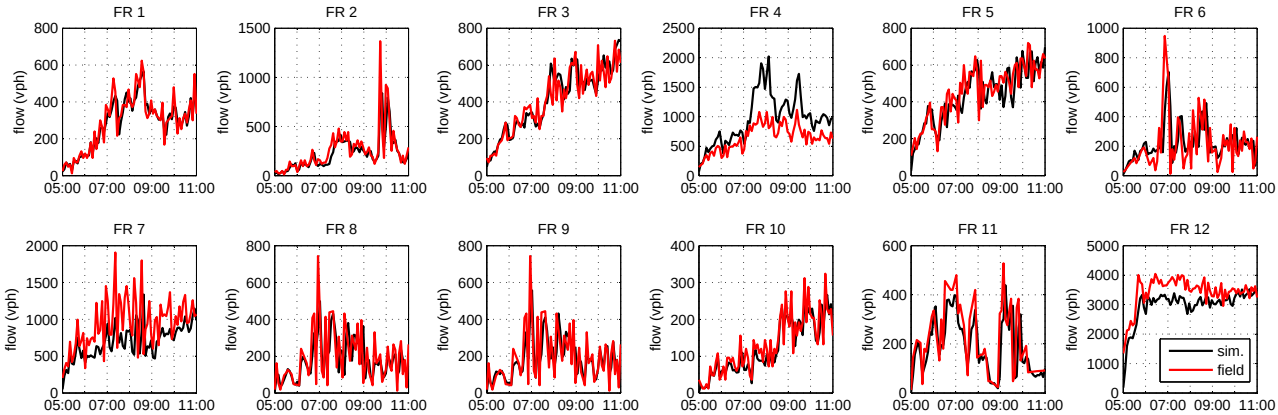


Figure 4.5: Simulated and measured flow at off-ramps

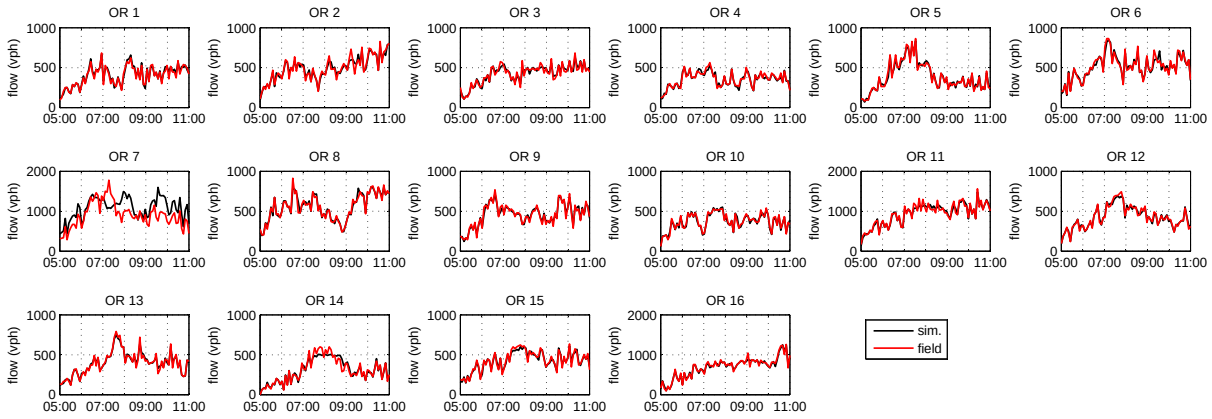


Figure 4.6: Simulated and measured flow at on-ramps

VDS ID	RMSPE	GEH	VDS ID	RMSPE	GEH
OR 1	6.64 %	100 %	FR 1	16.11 %	91.67 %
OR 2	7.56 %	97.22 %	FR 2	35.62 %	84.72 %
OR 3	7.96 %	98.61 %	FR 3	16.17 %	80.65 %
OR 4	6.17 %	100 %	FR 4	58.45 %	16.67 %
OR 5	6.35 %	100 %	FR 5	20.1 %	72.22 %
OR 6	6.07 %	97.22 %	FR 6	51.26 %	68.06 %
OR 7	30.34 %	29.17 %	FR 7	37.79 %	25.00 %
OR 8	5.93 %	98.1	FR 8	46.91 %	73.61 %
OR 9	5.85 %	100 %	FR 9	35.85 %	79.17 %
OR 10	5.88 %	100 %	FR 10	22.21 %	100 %
OR 11	6.94 %	100 %	FR 11	34.86 %	83.33 %
OR 12	5.03 %	100 %	FR 12	16.21 %	43.06
OR 13	6.34 %	100 %			
OR 14	12.16 %	95.83 %			
OR 15	6 %	100 %			
OR 16	5.73 %	97.22 %			

Table 4.6: Calibration error of on-ramps and off-ramps

4.5 Summary

This chapter demonstrates a complete procedure for constructing and calibrating a microsimulation model with multiple vehicle classes (car, HOV, and truck) and on-ramp metering control. The procedure includes gathering the freeway geometric information, processing field data from the PeMS database, imputation for missing field data, and estimation of split ratio and demand from field data. The procedure is applied to SR-99 Northbound, which presented several challenging features: 25% of the mainline data is problematic, 42% of the off-ramp data is missing, 16 metered on-ramps with and without HOV bypass lanes, mainline HOV lane, multiple vehicle classes and interacting bottlenecks. All these features are considered and included in this model. Traffic analysis results indicate that one bottleneck is induced by the diverging traffic of freeway interconnection, while another is caused by high demand merging traffic. Simulation results indicate that the calibrated microsimulation model can reproduce the field flow measurements and the on-ramp performance.

Chapter 5

Field Implementation of Coordinated Ramp Metering Control

5.1 Introduction

Ramp metering (RM) is the major method of freeway traffic control, a strategy for alleviating the effect of the capacity drop phenomenon, which negatively influences transportation system efficiency. In order to increase freeway efficiency, the outflow from an on-ramp into a freeway mainline is regulated to prevent mainline flow breakdown by storing surplus on-ramp vehicles in an on-ramp queue. Freeway capacity drop means that the congestion after a breakdown has an outflow lower than the free-flow capacity. When capacity drop occurs, the throughput of the freeway can be reduced by 5% to 20%, leading to a higher total travel time spent in the network.

As a result, a large amount of research has been conducted on designing effective ramp metering strategies. Different types of ramp metering strategies have been studied in field implementation tests in the past decades [8], [27]. Most Ramp Metering operations in California are fixed by time-of-day (TOD) or are locally responsive to occupancy measurements immediately upstream of the entrance ramp merge; the latter is called a local responsive ramp metering (LRRM) strategy. The locally responsive ramp metering strategy [46] adjusts the ramp metering rate of each on-ramp independently such that traffic flow at the entrance ramp merge area is regulated. However, the performance of locally responsive metering controller is limited, because traffic on each section of a freeway dynamically affects each other section. Downstream section flow depends on the demand flow from its upstream, and downstream congestion could back-propagate to the upstream. The global behavior of the network cannot be shared between local responsive controllers if there is no communication between them. This isolated control structure cannot coordinate available on-ramp storage space between freeway upstream and downstream, and, therefore, the overall mainstream capacity may be overloaded or underutilized.

In order to mitigate the drawbacks of LRRM, coordination control strategies have been

proposed in the literature to more effectively alleviate congestion. Coordinated ramp metering (CRM) is based on optimal control theory [81], [22], and it has been studied in analysis [66] and simulation [113] in several previous works, which have indicated some potential in reducing freeway congestion at recurrent bottleneck locations. Freeway corridor traffic flow is limited by bottleneck flow. If the section upstream of a bottleneck is congested, the bottleneck flow will drop well below its capacity. A logical approach to maximize flow at recurrent bottlenecks is to create a discharge section immediately upstream of the bottleneck.

This chapter presents a field implementation test and evaluation of a newly developed CRM algorithm [66]. This field implementation test is a collaborative project conducted by PATH research team [85] and Caltrans District 3 under the framework of the Caltrans District 3 traffic management center (TMC) system. The main tasks of the research include: (1) fine tuning CRM algorithms through simulation for California State Route 99 Northbound (SR-99N) morning peak traffic [113]; (2) implementing algorithm on PATH real-time control computers and integrating them with TMC ramp metering computers; (3) implementing and refining real-time traffic state parameter estimation; (4) conducting dry-runs with PATH computers interfaced with TMC computers; (5) progressively activating the CRM control; (6) evaluating the performance of CRM. These concepts need to be tested in the field to determine whether the projected benefits could be achieved in practice in California. If the results of field testing are favorable, they could provide the basis for future widespread adoption of CRM control strategies to further improve mobility and safety and reduce energy and emissions impacts of freeway congestion.

5.2 Test Site SR-99N and Available Data Description

The test site of the field implementation is on the SR-99N from Elk Grove Boulevard at absolute postmile 287.23 to the freeway interchange point of SR-99N and U.S. Route 50 (US-50) at absolute postmile 299.467, as shown in Figure 5.1. Figure 5.2 shows the typical speed contour for SR-99 Northbound, plotted from loop detector data on March 19, 2015. Two recurrent bottleneck locations are identifiable: Bottleneck 1, which is near the freeway interchange point of SR-99N and US-50 (PM 297.65); and Bottleneck 2, which is near Calvine Rd. (PM 290.76). The proposed CRM was deployed to on-ramps in the freeway segment on SR-99N between Calvine Road and the SR-50 interchange after 12th Avenue (abs. postmile 290.454 - 299.467). The simulation study of this test site has been conducted in Chapter 4. The field test involved activating the proposed CRM algorithm during morning peak traffic hours (6:00 AM - 9:00 AM) and evening peak traffic hours (3:00 PM - 6:00 PM). The test site is an 11-mile stretch of freeway that sustains heavy congestion during the morning and evening commutes. Morning congestion usually begins around 6:30 AM, peaks at 8:00 AM, and finally dissipates at around 10:00 AM. Evening congestion usually begins around 3:00 PM, peaks at 4:00 PM, and finally dissipates at around 7:00 PM. The leftmost lane is a high-occupancy vehicle (HOV) lane and spans the entire mainline of the test site. The HOV lane is reserved for vehicles carrying two or more passengers, and its operational time is

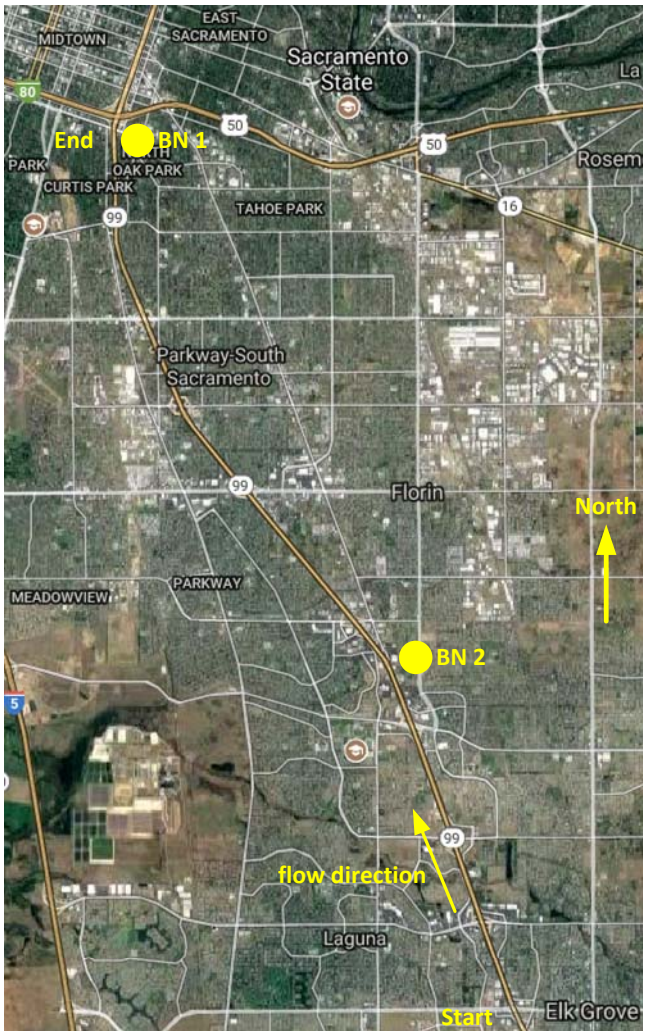


Figure 5.1: Satellite map of SR-99 N test site segment and its bottlenecks.

from 6:00 AM to 10:00 AM and from 3:00 PM to 7:00 PM. The whole test site has 16 on-ramps and 11 off-ramps. Each of these 16 on-ramps is equipped with ramp metering, which has a corresponding upstream mainline and downstream on-ramp detector station for traffic responsive control. Eleven of these 16 on-ramps have HOV lanes, which are individually metered by ramp metering. The network configuration and vehicle detection station (VDS) configuration is shown in Figure 4.1, where the numbered black boxes, red circles, and green circles label the indices of mainline VDS, off-ramp VDS, and on-ramp VDS, respectively. The dashed arrows in Figure 4.1 indicate that off-ramps 6, 7, 8, and 11 are missing. Those missing flow data were recovered by imputation [113], [72] using PeMS [86] historical data.

The field test implementation uses 30 seconds of raw data, which were directly collected from the loop detectors and 2070 traffic controllers at the test site. To monitor and control the test site, lane-by-lane mainline data (namely flow, speed, and occupancy) were collected

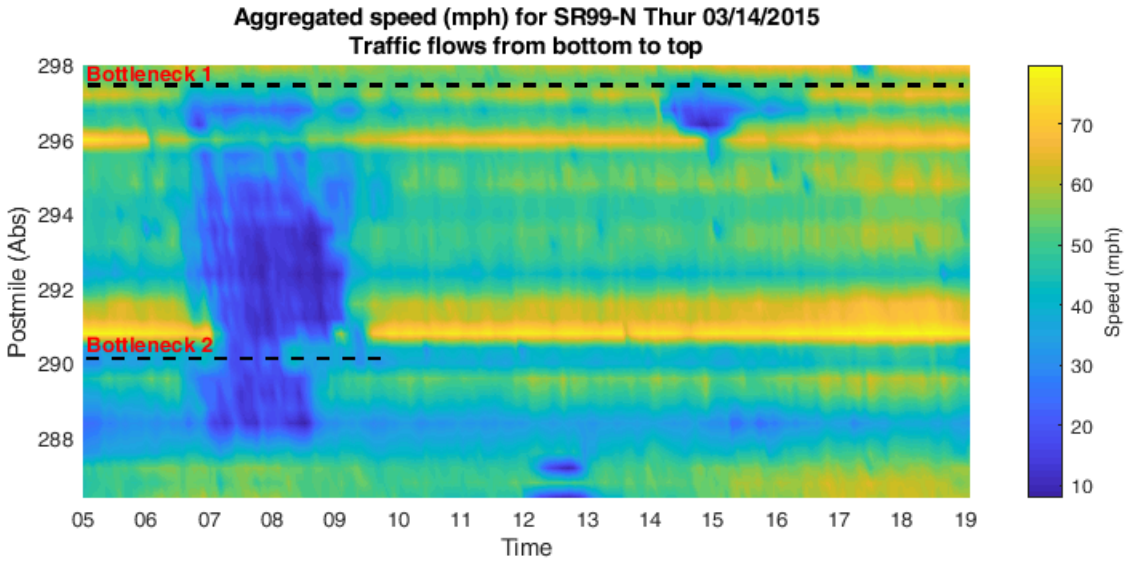


Figure 5.2: SR-99N speed contour, March 19, 2015.

from each loop detector. Further, lane-by-lane on-ramp and off-ramp flow and occupancy data are also collected from each loop detector. Mainline vehicle density were calculated by the speed-flow relation. These collected raw data were aggregated after outliers, abnormal data, and missing data were removed or imputed by a real-time data cleansing algorithm [64]. The aggregated data in this project were compared with data in a PeMS archive [86] in order to verify their correctness.

5.3 Coordinated Ramp Metering

The proposed CRM problem is formulated as a discrete time-optimal control problem with constraints on the decision variables over a finite horizon N_p, which is generally referred to as model predictive control (MPC). A cell transmission model (CTM) is used for the description of freeway traffic flow, constituting the modeling part of the MPC formulation. In order to increase the computation efficiency of solving optimization problems in real-time, we simplify the original nonlinear MPC problem such that it becomes a linear programming (LP) problem by assuming that average traffic speeds for each freeway section are available [66]. This assumption of having speed data is practical since good speed measurements can be obtained from dual loop traffic detector stations in the test field [64]. These dual loop detectors are connected to 2070 controllers under the universal ramp metering system (URMS) framework in TMC, and, therefore, other traffic state parameters (flow and occupancy) can also be obtained.

Freeway Model

Considering the macroscopic traffic model, a freeway is divided into N segments such that each segment i has at most one on-ramp and one off-ramp. The following traffic quantities are defined for each segment i at each time step k:

$\rho_i(k)$: Mainline density; number of vehicles in segment i at time step k.

$\bar{\rho}_i(k)$: Estimated/measured mainline density.

ρ_i^J : Jam density of segment i.

$f_i(k)$: Mainline flow (number of vehicles per time step) of vehicles leaving upstream segment i, moving to downstream segment $i+1$, at time step k.

$\bar{f}_i(k)$: Measured mainline flow.

F_i : Mainline flow capacity of segment i.

$w_i(k)$: Number of vehicles on the on-ramp corresponding to segment i, at time step k.

w_i^J : Jam density of the on-ramp corresponding to segment i.

$r_i(k)$: Metering flow rate; number of vehicles entering segment i through its on-ramp at time step k, determined by the controller in actuated on-ramps.

r_i^o : Maximum possible on-ramp flow for the on-ramp i.

$d_i(k)$: Estimated/measured on-ramp demand; flow of vehicles intending to enter the on-ramp belonging to segment i at time step k.

$s_i(k)$: Off-ramp flow; flow of vehicles that leave segment i through its off-ramp at time step k.

$v_i(k)$: Time mean speed of vehicles moving in segment i at time step k.

$u_i(k)$: Space mean speed of vehicles moving in segment i at time step k.

T : Sampling time or simulation time step size.

λ_i : Number of lanes in segment i.

L_i : Length of mainline segment i.

L_i^o : Queue capacity of on-ramp i; maximum number of vehicles that the on-ramp corresponding to segment i can accommodate.

N_p: Prediction horizon.

By the law of conservation, the dynamics of freeway mainlines are described by the evolution of mainline density $\rho_i(k)$ over time:

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{\lambda_i L_i}(f_{i-1}(k) + r_i(k) - f_i(k) - s_i(k)) \ . \tag{5.1}$$

Since traffic density estimated from a space mean speed $u_i(k)$ is more realistic [50], the traffic flow can be computed for each time step by

$$f_i(k) = \lambda_i \rho_i(k) u_i(k), \tag{5.2}$$

where space mean speed $u_i(k)$ is assumed to be given. Substituting equation (5.2) into (5.1) gives a linearized equation:

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{\lambda_i L_i}(\lambda_{i-1}\rho_{i-1}(k)u_{i-1}(k) + r_i(k) - \lambda_i \rho_i(k) u_i(k) - s_i(k)) \ . \tag{5.3}$$

Similarly, the evolution of on-ramp queue is described by the following conservation equation:

$$w_i(k+1) = w_i(k) + T(d_i(k) - r_i(k)) \ . \tag{5.4}$$

Supposing that there are n_i fixed sensors (loop detectors) on segment i and $\bar{v}_l(k)$ is measured speed from each sensor, the time mean speed is computed by

$$v_i(k) = \frac{1}{n_i} \sum_{l=1}^{n_i} \bar{v}_l(k) \tag{5.5}$$

Assuming stationary conditions, the space mean speed can be computed from local measurements $\bar{v}_l(k)$, using a harmonic mean of the measurements [50]:

$$u_i(k) = \frac{1}{\frac{1}{n_i}\sum_{l=1}^{n_i} \frac{1}{\bar{v}_l(k)}}. \tag{5.6}$$

Constraints

In reality, the controlled freeway is subjected to constraints for the maximum and minimum values of mainline density, on-ramp queue length, and ramp metering rate. These constraints are formulated as the following inequalities

$$0 \le w_i(k) \le L_i^o w_i^J \ , \tag{5.7}$$

$$0 \le r_i(k) \le \min\{d_i(k), r_i^o, \lambda_i(F_i - \bar{f}_{i-1}(k)), \lambda_i u_i(k)(\rho_i^J - \bar{\rho}_i(k))\} \ , \tag{5.8}$$

$$0 \le \rho_i(k) \le \min\{\rho_i^J, \phi(u_i(k))\} \ . \tag{5.9}$$

The first inequality constraint, in equation (5.7), is the entrance ramp queue length limit; the one in equation (5.8) represents the direct constraints on ramp metering rate, which is

the minimum of the four terms in the braces: the entrance ramp demand, entrance ramp capacity, and are space available in the mainline (the last two terms). $\lambda_i(F_i - \bar{f}_{i-1}(k))$ is related to the free-flow case, and $\lambda_i u_i(k)(\rho_i^J - \bar{\rho}_i(k))$ is related to the congestion case. The third inequality, in equation (5.9), is an indirect constraint on ramp metering rate through the density dynamics. The function $\phi(u_i(k))$ describes the speed versus density, which is obtained from an empirical study of traffic speed drop [18].

Objective Function

Under the framework of optimal control, controllers are typically designed to minimize or maximize a single objective function. For ramp metering control, total time spent (TTS) and total traveled distance (TTD) are two interesting objective functions. TTS is defined as

$$\mathrm{TTS}(k) = T \sum_{j=0}^{N_p-1} \sum_{i=1}^{N} L_i \lambda_i \rho_i(k+j) + \alpha_w T \sum_{j=0}^{N_p-1} \sum_{i=1}^{N} w_i(k+j), \tag{5.10}$$

where the first term of TTS is also called total travel time (TTT), the second term of TTS represents time delay due to on-ramp queue, and α_w is the on-ramp weighting parameter.

TTD is defined as

$$\mathrm{TTD}(k) = T \sum_{j=0}^{N_p-1} \sum_{i=1}^{N-1} L_i \lambda_i f_i(k+j) + T \sum_{j=0}^{N_p-1} L_N \lambda_N f_N(k+j). \tag{5.11}$$

For tractability reason, these two objective functions are combined into a single cost function

$$J = \mathrm{TTS} - \alpha \mathrm{TTD}, \tag{5.12}$$

where $\alpha > 0$ is a weighting parameter. Choosing the weighting parameters $\alpha_{\mathrm{TTD},N} \gg \alpha_{\mathrm{TTD},0} > 0$ emphasizes maximizing the flow on the most downstream segment N and equation (5.12) can be written as

$$\begin{aligned} J =& T \sum_{j=0}^{N_p-1} \sum_{i=1}^{N} L_i \lambda_i \rho_i(k+j) + \alpha_w T \sum_{j=0}^{N_p-1} \sum_{i=1}^{N} w_i(k+j) \\ & - \alpha_{\mathrm{TTD},0} T \sum_{j=0}^{N_p-1} \sum_{i=1}^{N-1} L_i \lambda_i f_i(k+j) - \alpha_{\mathrm{TTD},N} \sum_{j=0}^{N_p-1} L_N \lambda_N f_N(k+j). \end{aligned} \tag{5.13}$$

The reasons for choosing this objective function in equation (5.12) are as follows: in practice, TTS is related to vehicle-hour-traveled (VHT), and TTD is related to vehicle-miles-traveled (VMT). Both VHT and VMT are available in the PeMS system [86], and, therefore, it is convenient for any freeway user to evaluate the ramp metering control performance by accessing this open data base. In addition, minimizing TTS may discourage vehicles from

entering the freeway so that the mainline can have better flow when the mainline density is higher. Minimizing negative TTD is equivalent to maximizing (positive) TTD, which encourages vehicles get to enter the freeway. Therefore, to minimize the difference between TTS and TTD is to formulate the problem as a non-zero sum game. The overall effect of minimizing the objective function J is to minimize VHT and maximize VMT. In addition, since freeway efficiency (average speed) is defined as Q = VMT/VHT, minimizing the cost function J also lead to improved freeway efficiency Q.

Model Predictive Control Schemes

Taking together the freeway model, constraints, and the cost function, the CRM controller can be formulated according to the general formulation of an MPC controller (receding horizon predictive control) [12]. At each time step k, the CRM strategy is obtained by solving the following optimization problem

$$
\begin{aligned}
\underset{\rho,w,r}{\text{minimize}} \quad & J = \text{TTS} - \alpha \text{TTD} \\
\text{subject to} \quad & \rho_i(k+j+1) = \rho_i(k+j) + \frac{T}{\lambda_i L_i}(\lambda_{i-1}\rho_{i-1}(k+j)u_{i-1}(k+j) + r_i(k+j) \\
& - \lambda_i \rho_i(k+j)u_i(k+j) - s_i(k+j)), \\
& w_i(k+j+1) = w_i(k+j) + T(d_i(k+j) - r_i(k+j)), \\
& 0 \le \rho_i(k+j) \le \min\{\rho_i^J, \phi(u_i(k+j))\}, \\
& 0 \le w_i(k+j) \le L_i^o w_i^J, \\
& 0 \le r_i(k+j) \le \min\{d_i(k+j), r_i^o, \lambda_i(F_i - \bar{f}_{i-1}(k+j)), \lambda_i u_i(k+j)(\rho_i^J - \bar{\rho}_i(k+j))\}, \\
& \text{for} \;\; i = 1, \ldots, N, \;\; j = 0, \ldots, N_p - 1.
\end{aligned}
$$

Since this problem is a linear programing problem, it can be efficiently solved in real-time by the Simplex method [87].

On-Ramp Queue Override

The following on-ramp (entrance ramp) queue override scheme has been used jointly with the CRM algorithm. The queue detector is located about 15% distance to the upstream end of the entrance ramp. The schematic queue override rule is as follows:

1. If the occupancy of the queue detector is over 70%, then use the maximum lane ramp metering rate 950 (veh/hr) for 3 cycles (or 1.5 minutes).

2. If the occupancy of the queue detector continues to be higher than 70%, then this maximum lane ramp metering rate will remain.

Simulation Results

Before field test implementation, Aimsun microsimulation model is selected as a test platform for performance evaluation. The proposed CRM algorithm is implemented in the well calibrated model proposed in Chapter 4. The following system-wide performance indicators provided by Aimsun have been used for evaluation of the algorithm implemented in the simulation model: total travel time (TTT), total travel distance (TTD), total delay (TD) (obtained by deducting hypothetical free-flow travel times from simulated travel times), total number of stops (TNOS) (used as a system-wide performance parameter for traffic smoothness). By comparing the proposed CRM with the default RM, the improvement of an index x is calculated by

$$\Delta x = \frac{x_{\text{CRM}} - x_{\text{default RM}}}{x_{\text{default RM}}}. \tag{5.14}$$

In simulation, traffic data of four different days in 2015 is selected as model inputs, which contains all the on-ramps demand, off-ramp split ratio and freeway most upstream mainlines flow. The simulation model is executed during AM peak hours from 6:00 AM to 9:30 AM. Each day was run for 10 replications (random seeds). The 10 replications for each model day were also different. The results were then averaged over the 10 replications and it is shown in Table 5.1.

Date	Δ TTT	Δ TTD	Δ TD	Δ TNOS
March 18, 2015	-5.84%	0.78%	-12.53%	-2.42%
March 19, 2015	-6.29%	0.85%	-11.82%	-3.23%
September 17, 2015	-5.13%	0.77%	-11.73%	-2.68%
September 18, 2015	-5.33%	0.86%	-11.92%	-3.45%
Average	-5.65%	0.815%	-12%	-2.95%

Table 5.1: System-wide performance indicators changes with CRM control

5.4 Field Implementation

Concept of Operations

Figure 5.3 shows the overall system structure of the CRM system and signal flow of the system. The red arrow starting from the loop detector on the freeway to the PATH computer in the figure is the measurement of all available field data (flow, speed, occupancy). The blue arrow starting from the PATH computer to all cabinets (URMS controller in the field) in the figure is the calculated optimal ramp metering rate by the proposed algorithm. The yellow arrow in the figure starting from each cabinet (URMS) to its corresponding ramp metering traffic light is the on-ramp metering light control signal. The PATH CRM computer

is installed in the Caltrans District 3 TMC and directly links with its intranet for data acquisition, processing, traffic state parameter estimation, calculating the optimal ramp metering rate, and sending it to the corresponding on-ramp for activation. The PATH CRM computer collects traffic data and aggregates it every 30 seconds. The benefits of using this system structure are the following. The intranet connection with 2070 controllers in the field used fixed IP addresses. Such an implementation scheme is advantageous. First, it is simple and direct. Second, there is no middle system between the PATH CRM computer and the 2070 controllers in the field; therefore, third party support is not necessary. Thirdly, the PATH computer can access all the raw field data unchanged by any middle system; therefore, the data are trustworthy. Fourth, such a direct link practically avoids any delays and data passing errors caused by middle systems.

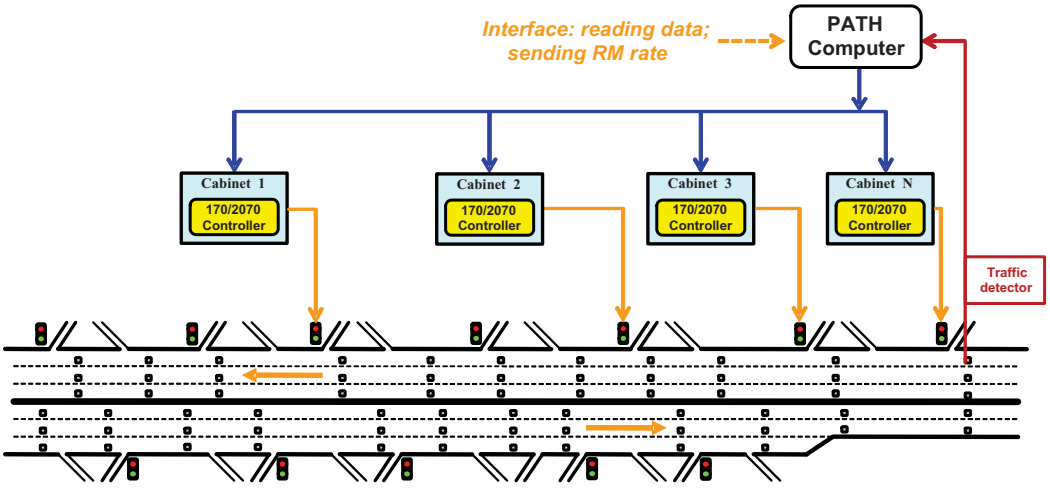


Figure 5.3: Direct interface between TMC ramp metering computer and CRM controllers; PATH computer is for data processing and calculation of CRM rate.

Progressive Implementation

Before changing the original fixed-time local responsive ramp metering strategy into the proposed CRM algorithm, the implemented system runs in a dry-run mode. In this mode, the computer sends feedback signals (from sensor measurements) from the system to the controller such that the controller generates a control signal but without implementing the control signal to the system. By investigating these control signals, one can check the correctness of control signals, and the parameters in the system can be fine-tuned. The dry-run phase of this project occurred in the week of September 19-25, 2016. During the dry-run, calculated CRM rates were saved to files instead of being sent to 2070 controllers for activation. Those saved files were carefully checked to make sure every part of the system worked correctly and robustly in the sense that, even if there was some fault in the loop detector data, historical data from the same time of a day would be used as a backup, and,

therefore, sensor faults would not significantly affect the CRM calculation. In case there was a problem in the PATH computer, the 2070 controller in the field would automatically activate the default LRRM, deactivating CRM. On the day when the CRM was switched on, the project team monitored the CRM system, adjusted the parameters in the algorithm, and observed the traffic through D3 freeway traffic video systems and Google traffic maps. The comparison of CRM rates with default LRRM rates were studied.

The formal tests began on September 26, 2016, although some minor adjustments were still made in the second week. Then, the algorithm was finalized and extensive data collection began. The project team keep recording and monitoring traffic data in the field, included comparison of ramp metering rates of CRM and original local responsive ramp metering for both AM and PM peak hours, as well as some other freeway traffic parameters.

Monitoring of CRM Rate

To make sure the CRM algorithms were executed correctly, the project team monitored all traffic data obtained from 2070 controllers during the field test. Information that can be obtained from the 2070 controller include: onramp name, machine time, field RM ID, control scheme currently activated (i.e., LRRM or CRM), and cycle count. This information can be used to tell if the CRM algorithm is activated and identify the problem if is not activated. Additionally, LRRM and CRM rates were compared for everyday during the tests. Since the project team only deployed the proposed CRM algorithm in the 11 downstream on-ramps (the first 5 on-ramps used the original LRRM strategy), it was sufficient to monitor the downstream 11 on-ramps. Figure 5.4 and Figure 5.5 presents an example comparison of ramp metering rates in morning and evening peak traffic, respectively. The names of those 11 downstream on-ramps are listed in Table 5.2. The ramp metering rates for LRRM (red lines) and CRM (blue lines) control strategies are quite different for AM peak hours. The results indicates that the proposed CRM can respond to traffic demand (green lines), since the CRM changes with time and demand differences, whereas most of the LRRMs give constant metering rates almost all the time. The LRRM metering gives a high ramp metering value in ramps 14 and 15 from 6:00 AM to 6:30 AM, since the demand is low during that time.

Monitoring of Queue Length

The proposed CRM algorithm changes the ramp metering light and it also influence the queue dynamics at each on-ramp. Since the queue detector in the test site is not available, the actual queue length during the field test cannot be measured directly. Although the queue length can be estimated by other information around the on-ramp: on-ramp flow, demand, and its adjacent mainline flow, the accuracy of queue length estimation is very limited. In order to overcome this equipment limitation, Google Map is used for monitoring the queue condition at each on-ramp. Besides, Google Map also provides freeway incident and accident information in real-time, which help us understand if there is an event influence

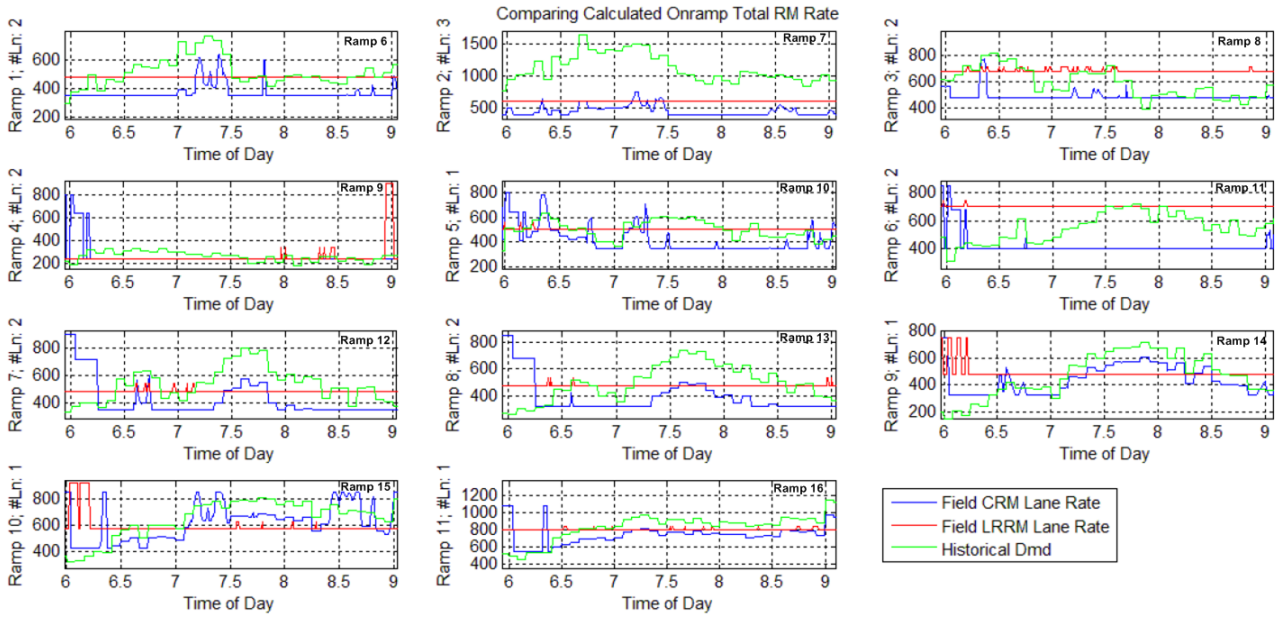


Figure 5.4: Comparison of LRRM and CRM rates for AM peak hours on Wednesday, 12 October, 2016.

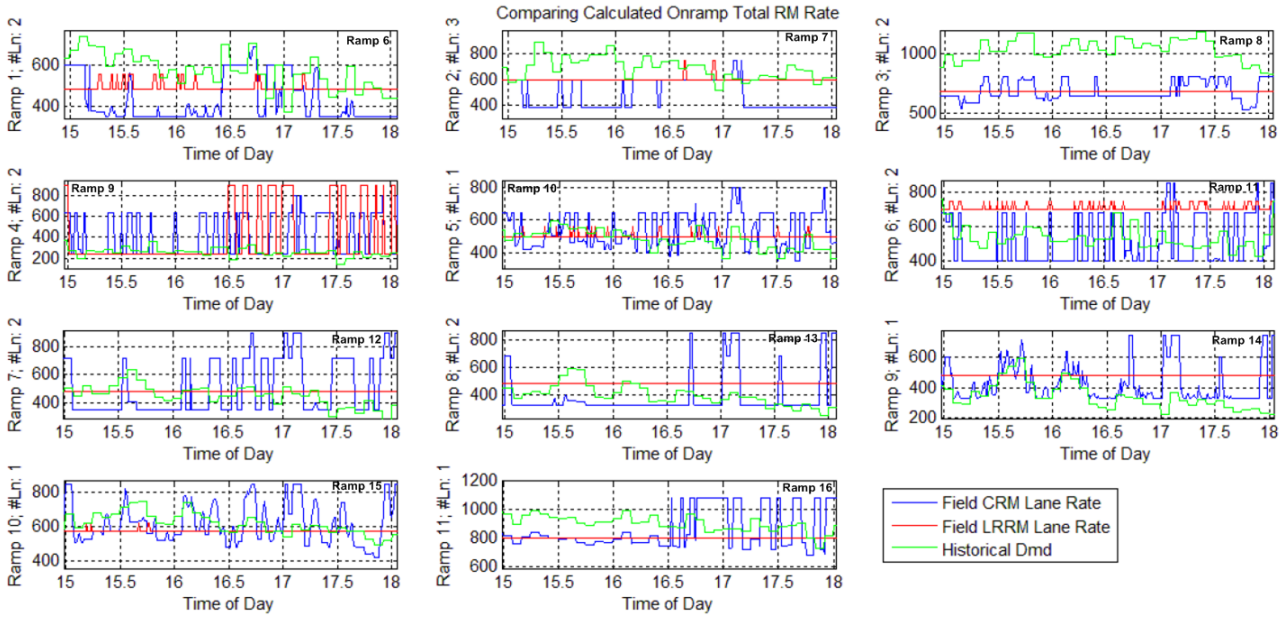


Figure 5.5: Comparison of LRRM and CRM rates for PM peak hours on Wednesday, 12 October, 2016.

the field test. The observation was made from 6:00 AM to 9:00 AM every 20 minutes to 30 minutes during the field test. Figure 5.6, Figure 5.7, Figure 5.8, Figure 5.9, Figure 5.10, and Figure 5.11 are queue monitoring of 11 downstram on-ramps at 7:22 AM on Wednesday, 12 October, 2016. The other monitoring of queue length during morning peak hours by Google Map on Wednesday, 12 October, 2016 is provided in Appendix A. The observation of queue length by Google Map indicates that the queue at each on-ramp will not split back to its adjacent local street and arterial. The on-ramp storage is used without excess its capacity.

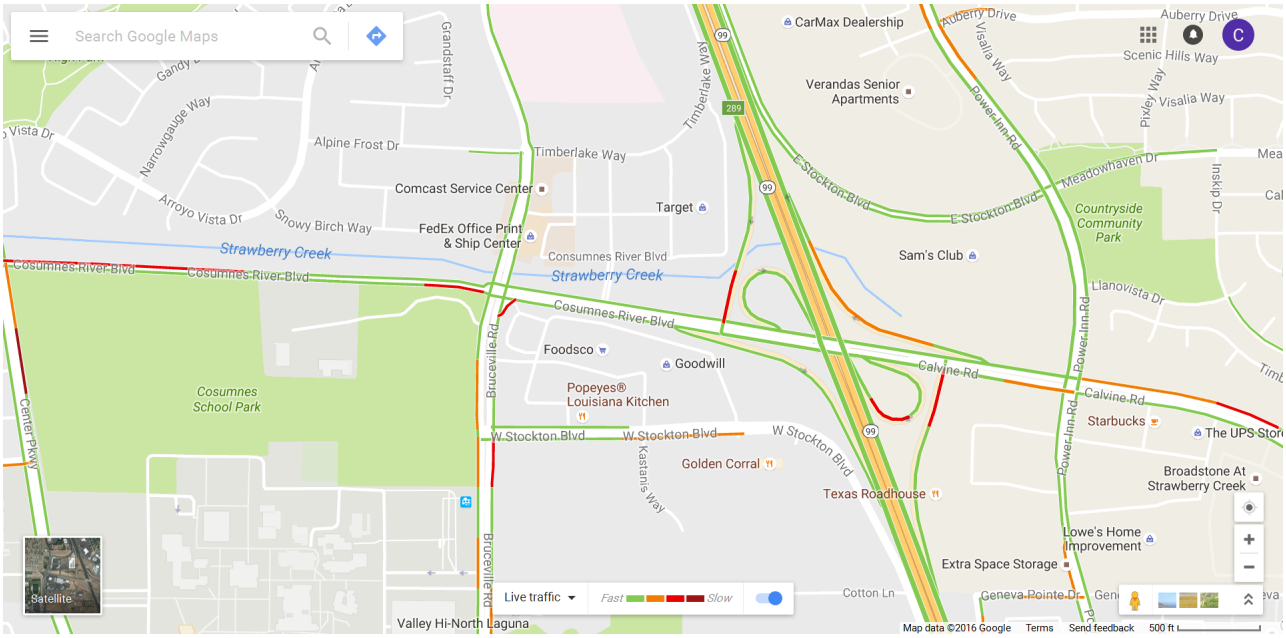


Figure 5.6: Monitoring of queue length near EB Calvine Road on-ramp (OR6) and WB Calvine Road on-ramp (OR 7) at 7:22 AM on Wednesday, 12 October, 2016

5.5 Results and Discussion

The data for performance evaluation field tests were obtained from PeMS [86]. The data for the stretch of test site SR-99N from postmile 280 to postmile 300 are used. The sampling time of performance indexes VMT, VHT, and Q is the minimal sampling time provided by PeMS in hours. Data from before the field test are taken from weekdays from the first week in October through the first week in November, 2015. Field test data are taken from weekdays from the first week in October to the first week in November, 2016. The AM ramp metering activation time is from 6:00 AM to 9:00 AM. The PM ramp metering activation time is from 3:00 PM to 6:00 PM. Note that the field test started from Wednesday, September 19, 2016. The traffic data from September 19, 2016, to September 30, 2016, are omitted, since the project team was adjusting system parameters during the beginning of the field test. After

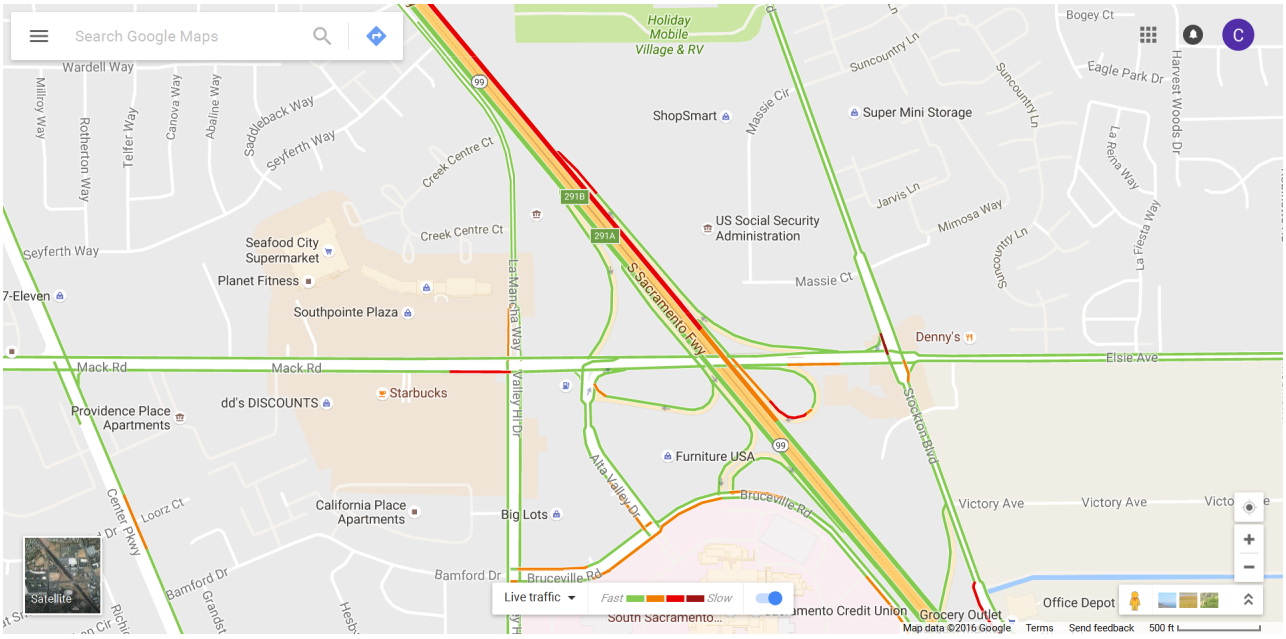


Figure 5.7: Monitoring of queue length near EB Mack Road on-ramp (OR 8) and WB Mack Road on-ramp (OR 9) at 7:22 AM on Wednesday, 12 October, 2016

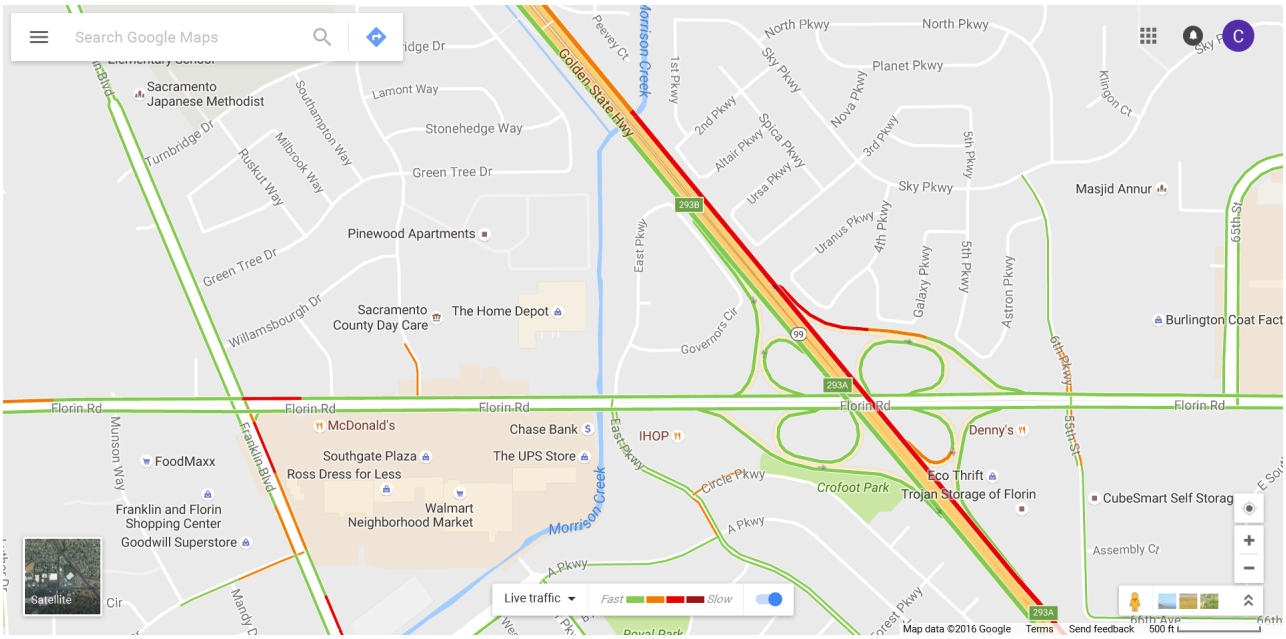


Figure 5.8: Monitoring of queue length near EB Florin Road on-ramp (OR 10) and WB Florin Road on-ramp (OR 11) at 7:22 AM on Wednesday, 12 October, 2016

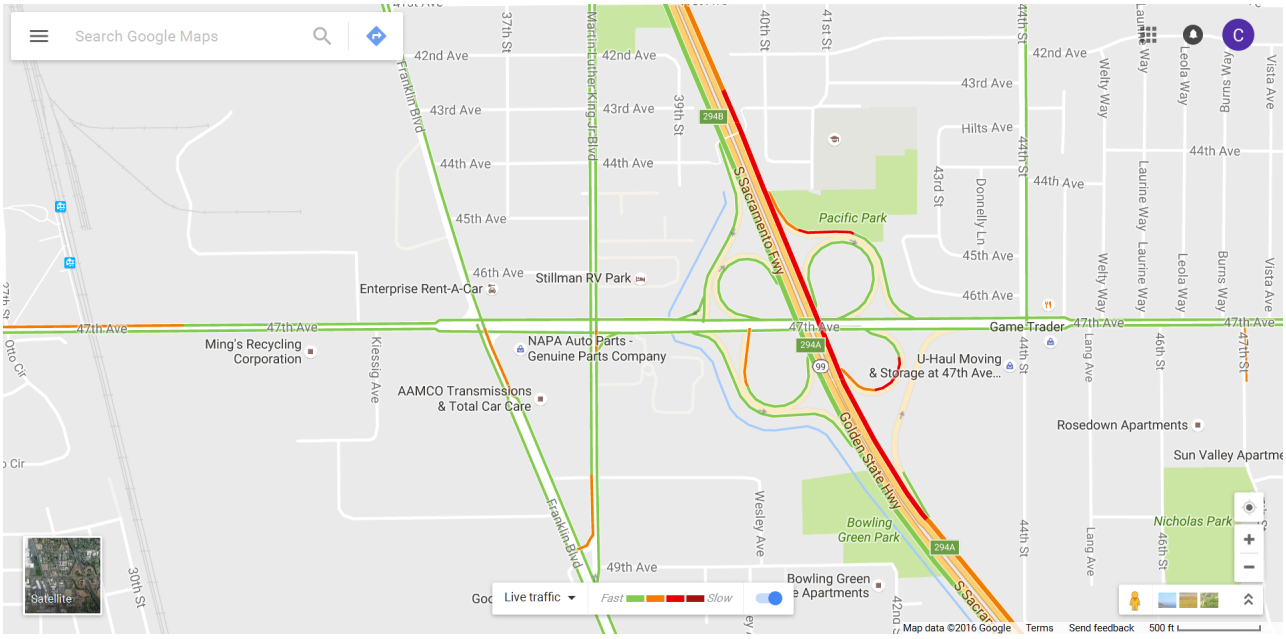


Figure 5.9: Monitoring of queue length near EB 47th Ave on-ramp (OR 12) and WB 47th Ave on-ramp (OR 13) at 7:22 AM on Wednesday, 12 October, 2016

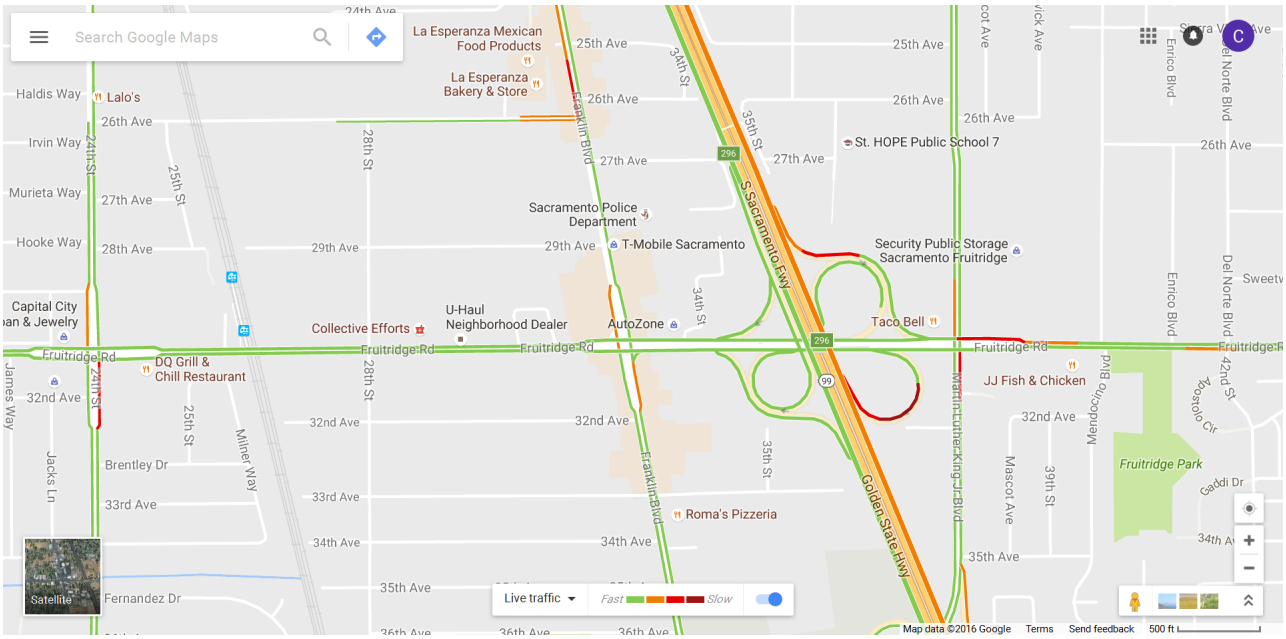


Figure 5.10: Monitoring of queue length near EB Fruitridge Road on-ramp (OR 14) and WB Fruitridge Road on-ramp (OR 15) at 7:22 AM on Wednesday, 12 October, 2016

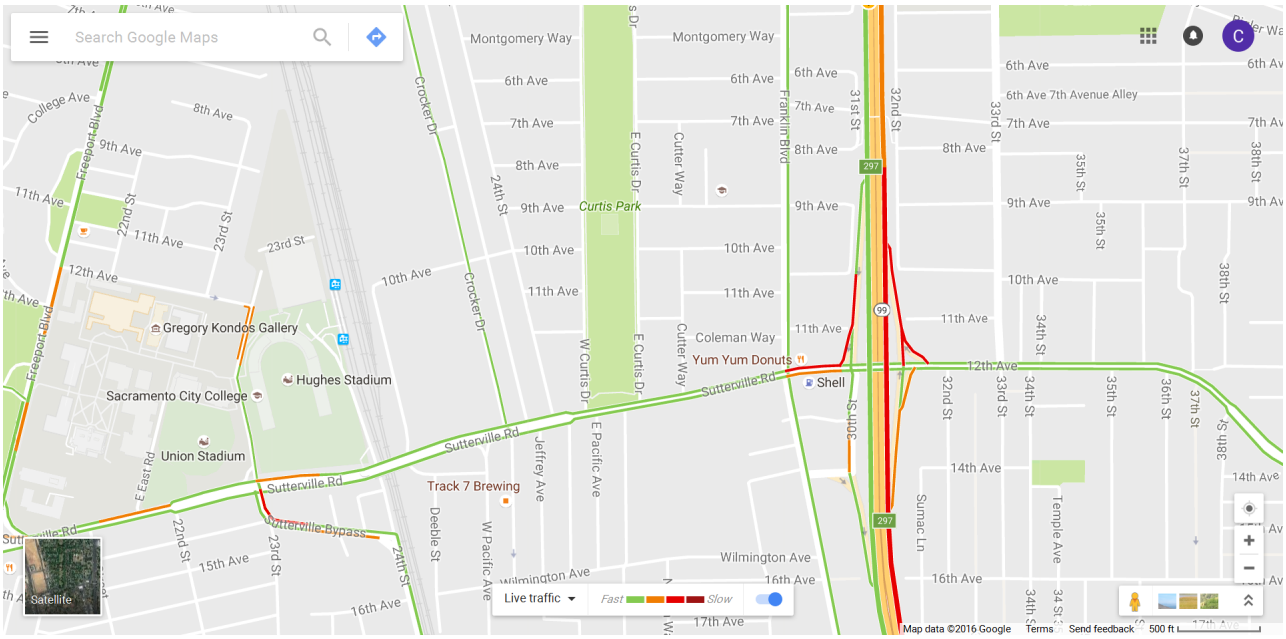


Figure 5.11: Monitoring of queue length near 12th Ave on-ramp (OR 16) at 7:22 AM on Wednesday, 12 October, 2016

the duration of system tuning, the traffic characteristics become more representative, making those data meaningful for analysis. The performance indexes are compared for the same day of the week for the weeks investigated; for example, Tuesdays in 2015 are compared with Tuesdays in 2016. This is a reasonable comparison because commuter traffic patterns are similar for the same weekdays.

Freeway System Performance Indexes

Coordinated ramp metering has been introduced mainly for reducing congestion by regulating the number of vehicles entering the freeway mainline from the on-ramp. There are several congestion related indicators: freeway efficiency (average speed) [15], average travel time, duration of congestion, fuel consumption and emissions index [83]. Since the goal of this project is to evaluate the influence of ramp metering control on a freeway system, we only consider TTD, TTT/TTS, and mean speed (MS). These traffic indices are equivalent to VMT, VHT and Q, respectively, in the PeMS system [86]. The calculation of these performance indices in PeMS are as follows.

VMT is the sum of distance (in miles) traveled by each vehicle on the given section of freeway over a given time period. VMT is equivalent to TTD. Consider a freeway is partitioned into N segments with length L_i for the i-th segment where each segment contains

at least one loop detector. VMT is computed by

$$\text{VMT} = \sum_{i=1}^{N} \text{VMT}_i, \tag{5.15}$$

where $\text{VMT}_i(k) = \bar{f}_i(k) L_i$ and $\bar{f}_i(k)$ is flow measurement at segment i.

VHT is the sum of all trip times (in hours) spent by each vehicle on the given section of freeway over a given time period. VHT is equivalent to TTT. VHT is computed by

$$\text{VHT} = \sum_{i=1}^{N} \text{VHT}_i, \tag{5.16}$$

where $\text{VHT}_i(k) = \text{VMT}_i(k)/v_i(k)$ and $v_i(k)$ is the speed at the i-th segment.

Freeway efficiency (mean speed) is expressed in miles per hour (MPH) and is defined as

$$\text{Q} = \frac{\text{VMT}}{\text{VHT}}. \tag{5.17}$$

From the definition of Q in equation (5.17), VMT is in the numerator of the Q value and VHT is in the denominator, so increasing VMT or decreasing VHT can make Q increase, which is consistent with the control objective: maximize TTD or minimize TTS. Therefore, higher Q values not only indicate better control performance, but also better freeway efficiency.

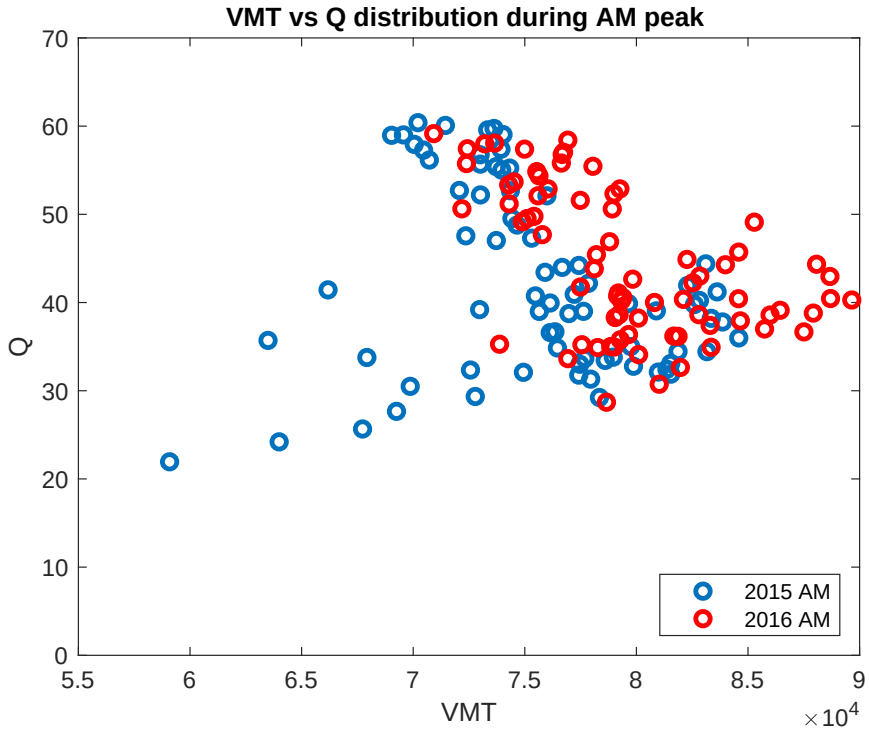
There is another way to interpret the freeway efficiency Q. Since the unit of Q is the same as flow over vehicle density (f/ρ), freeway efficiency can also be interpreted as the mean speed of all trips of the freeway during a period of time. Higher Q values indicate that the drivers on the freeway gain higher speed on average. Therefore, high Q values indicate high freeway efficiency. In addition, from a traffic engineering perspective, higher VMT values indicate that the freeway can be used by more drivers. Lower VHT values indicate the driver can spend less time travelling on the freeway. Increasing VMT or decreasing VHT can increase Q, which is equivalent to increasing freeway usage or to reducing travel time. Therefore, Q is an index of freeway efficiency for both traffic engineers and drivers.

Field Test Results

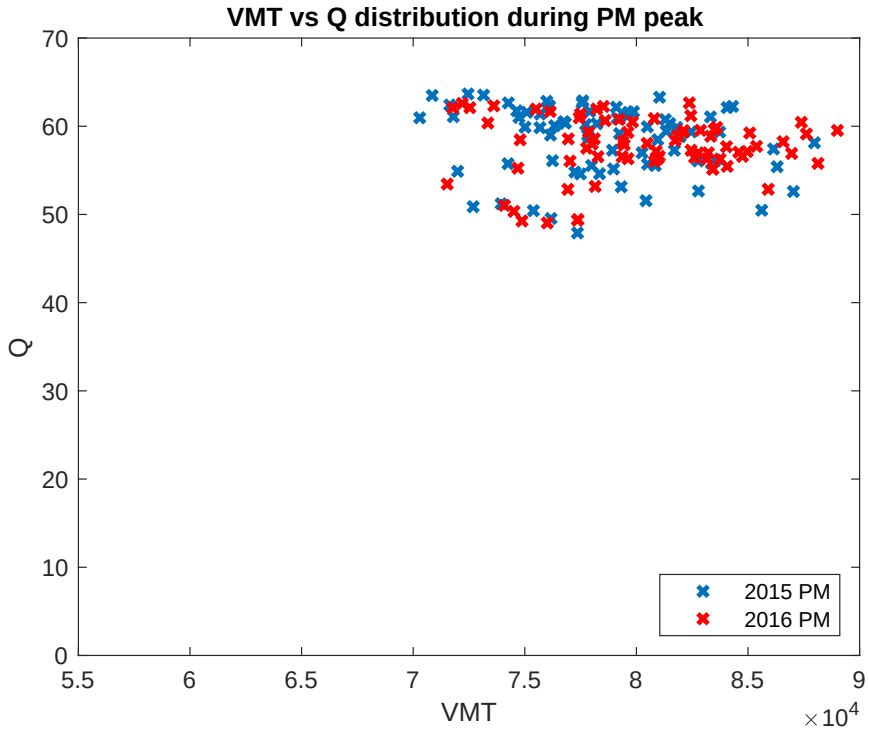
By comparing the traffic performance indexes before and after the field test, the improvement resulting from the proposed CRM strategy can be observed. The improvement of an index x is computed by

$$\Delta x = \frac{x_{\text{after test}} - x_{\text{before test}}}{x_{\text{before test}}}. \tag{5.18}$$

Figure 5.12 shows the VMT versus Q distribution, where circles represent the data from morning traffic and crosses represent data from evening traffic. The figure shows that the



(a)



(b)

Figure 5.12: VMT versus Q distribution, blue and red colors are traffic data in 2015 and 2016, respectively: (a) during morning peak (b) during evening peak.

circles are more scattered than the crosses, indicating that the VMT values have larger variation during the morning traffic than during the evening traffic. Red circles represent 2016 field test data, while blue circles represent 2015 pretest field data. Compared with the blue circle data cluster, the red circle cluster lies on the right of the VMT axis and above the Q axis. This comparison indicates the improvement of freeway efficiency indexes Q and VMT, since more red circles move toward positive Q and positive VMT directions, and it means the field test increased the freeway efficiency (average speed) Q by increasing VMT (usage of freeway or increasing demand). Conversely, the scatter of crosses in the figure is more concentrated than that of circles, which means the evening traffic did not change much after the CRM control. Therefore, the CRM control increased both the freeway efficiency and usage during morning traffic more significantly than it did during evening traffic.

The VHT improvement can also be observed by plotting VHT versus Q for both before and after the field test, as shown in Figure 5.13. In this figure, circles represent the data from morning traffic, and crosses represent data from evening traffic. The circles are more scattered than the crosses, indicating that the VHT values have larger variation during the morning traffic than during the evening traffic (the variation of travel time in AM is larger than in PM). Again, compared with the blue circle data cluster, the red circle cluster slightly moves to the left of the VHT axis and above the Q axis. This confirms the finding that the CRM control resulted in greater improvement in the morning than in the evening.

A summary of the PeMS data comparing avergae VMT, VHT, and Q before and after the field test is contained in Table 5.3, where **bold text** indicates improvement of the networks performance and *italic text* means deterioration of performance.

Morning ramp metering performance is summarized as follows:

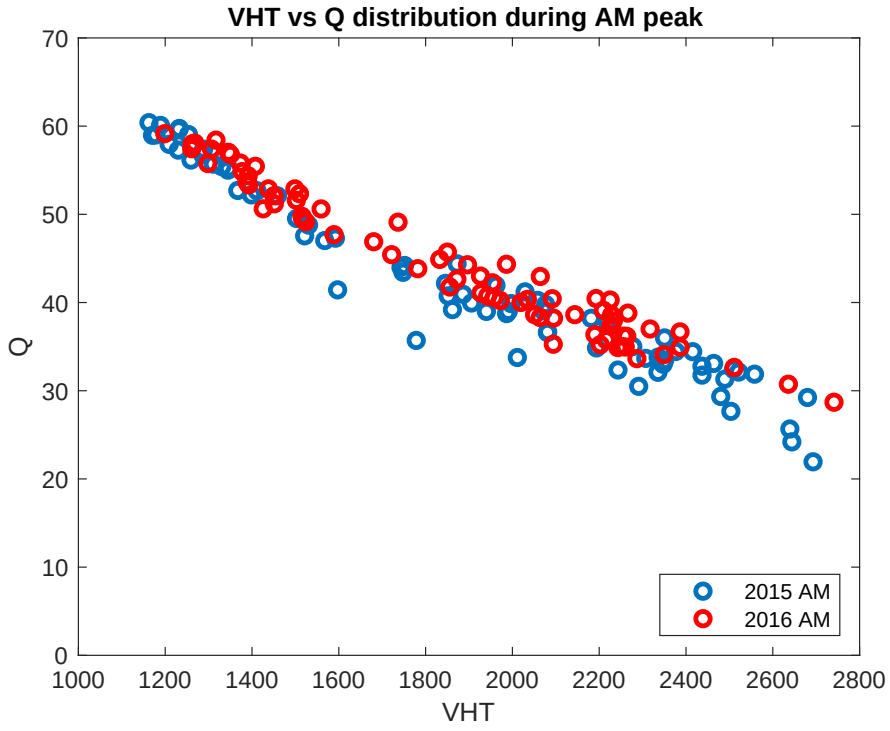
- VMT increased by 5.39% on average.
- VHT decreased by 1.64% on average.
- Q increased by 7.25% on average.

Since VMT and Q increased, we conclude that CRM improved the traffic during morning peak hours.

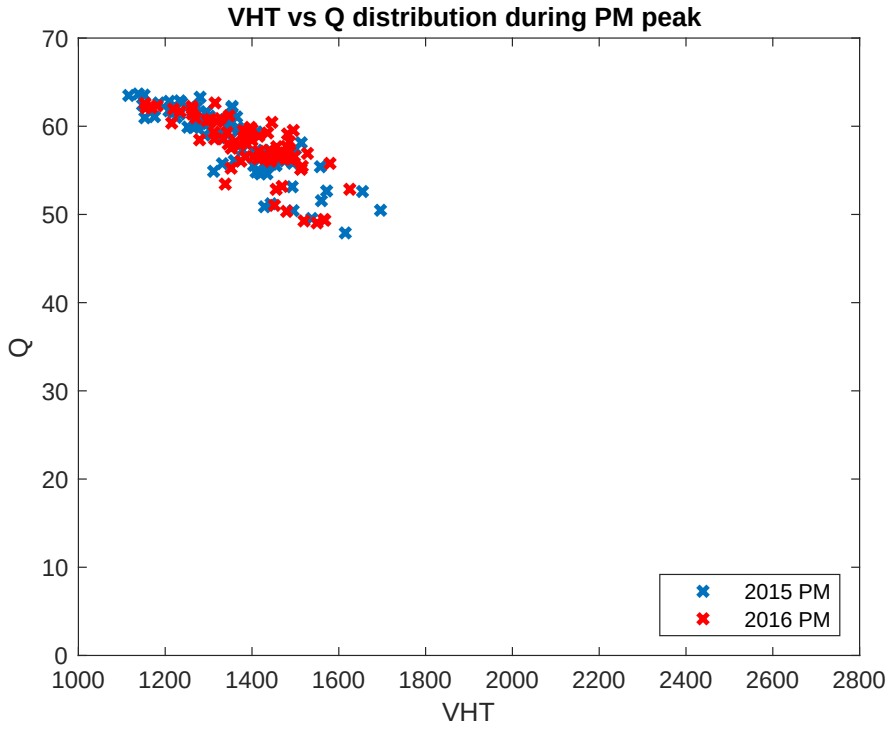
Evening ramp metering performance is summarized as follows:

- VMT increased by 2.56% on average.
- VHT increased by 3.04% on average.
- Q decreased by 0.44% on average.

Since the change in both VMT and Q was marginal, we conclude that CRM did not improve traffic during the evening peak hours.



(a)



(b)

Figure 5.13: VHT versus Q distribution, blue and red colors are traffic data in 2015 and 2016, respectively: (a) during morning peak (b) during evening peak.

The performance evaluation over the five weeks of data showed that freeway efficiency Q was increased by 7.25% on average for AM peak hours (from 6:00 AM to 9:00 AM) which usually had congested traffic. For PM peak hours (from 3:00 PM to 6:00 PM), freeway efficiency Q was decreased by 0.44% on average, which was within the statistical error margin. Therefore, the CRM algorithm could not improve PM traffic. The obvious reason for a significant lack of improvement during the evening commute is that the traffic was not congested most of the time in the PM peak hours, making it difficult for the CRM algorithm to improve it. This observation suggests that the CRM algorithm is most effective during congested traffic caused by high demand.

5.6 Summary

In this chapter, the results of a field test of CRM based on MPC is presented. The stretch of SR-99N between Elk Grove Street and the interchange with SR-50 was used as the test site. Previously, LRRM was used for the operation along that corridor in peak hours, and LRRM was still used for the first five upstream on-ramps in this project. The proposed CRM was implemented for the 11 downstream ramp meters. CRM is essentially different from LRRM in the sense that LRRM determines on-ramp metering rates based only on local mainline occupancy/flow data from its immediate upstream detector, whereas CRM determines ramp metering rates by using mainline occupancy/flow/speed data of the whole freeway and the demands of all on-ramps. Mathematically, centralized optimization is implemented for this small freeway using an efficient linear programing solver. Physically, the implemented algorithm controls the SR-99N corridor as a long discharging section in the sense that the downstream should not be more congested than the upstream traffic on average. We believe that this is the best strategy to alleviate traffic congestion on freeway.

After the control system was implemented, a progressive test procedure was conducted for the field tests. First, a dry-run was conducted for the first 2-3 weeks to verify that the overall system was running well. Then, after 5 weeks of tests and data collection, PeMS hourly VHT and VMT data were used for performance evaluation. It is noted that PeMS data are completely independent from the data in PATH computer obtained directly from the 2070 controllers in the field. By using PeMS data, the project team intended to obtain objective performance results to the extent possible. To address demand fluctuation, freeway efficiency (VMT to VHT ratio) was used as the performance parameter, which could be understood as the average speed. We believed that this ratio could reasonably accommodate traffic demand fluctuation. The aggregated data over five weeks for VMT/VHT increased by 7.25% for morning peak hours, indicating improvement in congested traffic. For evening peak hours, the CRM algorithm did not improve traffic, since the traffic is not heavily congested during that period. This suggests that the algorithm is most effective during congested traffic caused by high demand.

Index	Name	Control strategy	Number of Lanes	Number of HOV lanes
OR 1	Elk Grove Blvd	LRRM	3	1
OR 2	EB Laguna Blvd	LRRM	1	0
OR 3	WB Laguna Blvd	LRRM	1	0
OR 4	EB Sheldon Rd	LRRM	2	1
OR 5	WB Sheldon Rd	LRRM	3	1
OR 6	EB Calvine Rd	CRM	2	1
OR 7	WB Calvine Rd	CRM	3	1
OR 8	EB Mack Rd	CRM	2	1
OR 9	WB Mack Rd	CRM	2	1
OR 10	EB Florin Rd	CRM	1	0
OR 11	WB Florin Rd	CRM	2	1
OR 12	EB 47th Ave	CRM	2	1
OR 13	WB 47th Ave	CRM	2	1
OR 14	EB Fruitridge Rd	CRM	1	0
OR 15	WB Fruitridge Rd	CRM	1	0
OR 16	12th Ave	CRM	1	0

Table 5.2: List of on-ramp names

	6-7 AM	7-8 AM	8-9 AM	3-4 PM	4-5 PM	5-6 PM
2015 VMT	80118.58	74488.19	71804.62	78513.72	75687.66	69856.27
2016 VMT	83900.23	79286.52	75408.84	80324.92	76949.2	72439.83
2015 VHT	2324.12	2020.70	1366.42	1331.05	1305.34	1180.34
2016 VHT	2210.69	1973.29	1398.04	1410.57	1317.57	1206.86
2015 Q	34.47	36.86	52.54945	58.99	57.98	59.18
2016 Q	37.95	40.18	53.93897	56.95	58.40	60.02
Δ VMT	**4.72**%	**6.442**%	**5.019**%	**2.307**%	**1.667**%	**3.698**%
Δ VHT	**−4.881**%	**−2.346**%	*2.314*%	*5.974*%	*0.937*%	*2.247*%
Δ Q	**10.093**%	**8.999**%	**2.644**%	*−3.460*%	**0.723**%	**1.420**%

Table 5.3: Summary of both AM and PM hourly performance comparison

Chapter 6

Field Implementation of Variable Speed Advisory Control

6.1 Introduction

Variable Speed Advisory (VSA) control seeks to improve freeway efficiency by deferring the onset of congestion or increasing the effective throughput particularly at congested bottlenecks. Early research regarding Variable Speed Limit (VSL) focused on the concept of speed homogenization [100] [114]. The research shows that harmonizing the speed along the freeway and reducing speed differences in different lanes and for lane-changing can improve freeway efficiency [102]. In practice, VSL/VSA systems have been more widely deployed in Europe [90] than in the United States [52] [62]. VSL posts speed limits that drivers must obey, while VSA posts recommended driving speeds that are not legally enforced. Therefore, driver's willingness to comply with these two types of methods of control could be different and produce different levels of driver compliance rates. By the type of control algorithm, VSL/VSA can be divided into two categories: rule-based [55] [56] and model-based control [109] [108]. Rule-based VSA system uses pre-defined logic rules and preselect parameter thresholds to create real-time traffic control, while model-based VSA system uses a pre-established optimization model with traffic measurement to obtain optimal control action. Rule-based VSA strategies have been deployed in the United States, namely in Washington [110], Minnesota [69] [53], Oregon [92] and Missouri [48], and in European countries such as the United Kingdom [37], Germany [6], and the Netherlands [40]. Few model-based VSAs have been deployed in the real-world. One example is the SPECIALIST algorithm (SPEed ControllIng ALgorIthm using Shock wave Theory) and it have performed a field test on a 14 km long stretch on the Dutch A12 freeway [38]. This field test example was designed primarily to address recurrent bottlenecks or moving jams, in which an increase of throughput typically ranges from 5% to 15%, and to reduce of rear-end collisions.

In California, United States, most of the VSA deployments were originally designed for warning drivers of weather/road related hazardous conditions, visibility conditions, work

zones, incidents, and on-ramp/off-ramp/lane closures. A temporary VSA deployment in these situations is used for managing non-recurring congestion and increasing freeway safety. However, there have been a few temporary VSA deployments for recurring congestion in California. Developing an efficient variable speed advisory system that can regulate the traffic speed levels under the dynamically changing traffic conditions is necessary for improving freeway safety and alleviating recurrent congestion.

This chapter presents the field implementation test of variable speed advisory (VSA) for bottleneck flow maximization based on speed and occupancy measurement at the California State Route 78 Eastbound (SR-78E) corridor in San Diego. The procedure of design VSA control rules and hardware implementation is explained in this chapter. Field test results is evaluated by PeMS freeway performance indicators: VMT, VHT, and the VMT to VHT ratio. It is found that the proposed VSA control rules can be deployed for heavily congested traffic to alleviate congestion.

6.2 Test Site SR-78E and Available Data Description

The test site of the variable speed advisory (VSA) field implementation is on the SR-78E from Vista Village Drive in the City of Vista at absolute postmile 6.32 to the freeway interchange point of SR-78E, and U.S. Route 15 (US-15) in the city of Escondido, at absolute postmile 17.73, as shown in Figure 6.1. This test segment is a three-lane freeway with a posted speed limit of 65 mph and it has 10 on-ramps and 10 off-ramps. The available vehicle detector stations (VDS) are shown in Figure 6.2. A fixed message sign (FMS) displaying "FOLLOW ADVISORY SPEED" was placed at the starting point of the test site to instruct drivers to obey the speed posted by the downstream VSA. The posted speed on VSA during morning and evening peak hours is recommended to drivers but not enforced. The VDS are installed on the freeway mainline, on-ramps, and off-ramps to collect traffic data, namely volume, speed, and occupancy, every 30 seconds. This data is sent to PeMS for archiving and to our system for monitoring and controlling the freeway.

Bottleneck Identification

The scope of this field test is limited to reducing recurrent bottlenecks during morning and evening peak hours. From daily observation during weekdays, the morning peak hours range from 6:00 AM to 9:00 AM and the evening peak hours range from 2:00 PM to 19:00 PM. During peak hours, the speed drops from 60 mph to as low as 15 mph after the onset of the congestion. Figure 6.3 shows the speed contour for SR-78 Eastbound, plotted from loop detector data on March 14, 2018. Two recurrent bottleneck locations are identifiable: Bottleneck 1, which is near San Marcos Blvd. (PM 12.27); and Bottleneck 2, which is near the freeway interchange point of SR-78E and US-15 (PM 16.6). Bottleneck 2 (the downstream bottleneck) is caused by diverging traffic from SR-78E to US-15 NB and US-15 SB. This congestion may propagate back to midstream and activates bottleneck 1 at

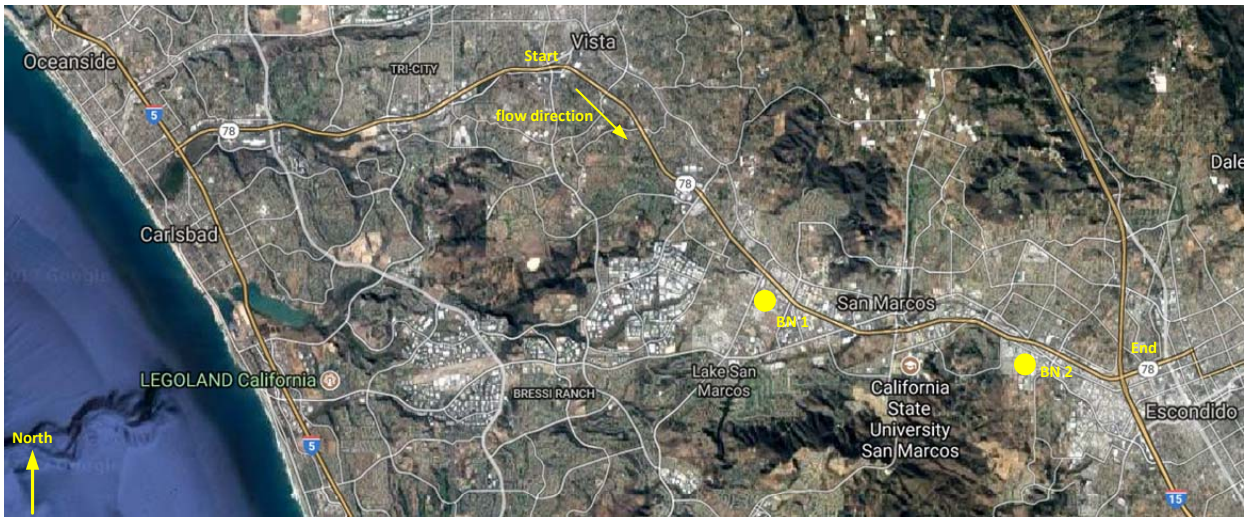


Figure 6.1: Satellite map of SR-78 E test site segment and its bottlenecks.

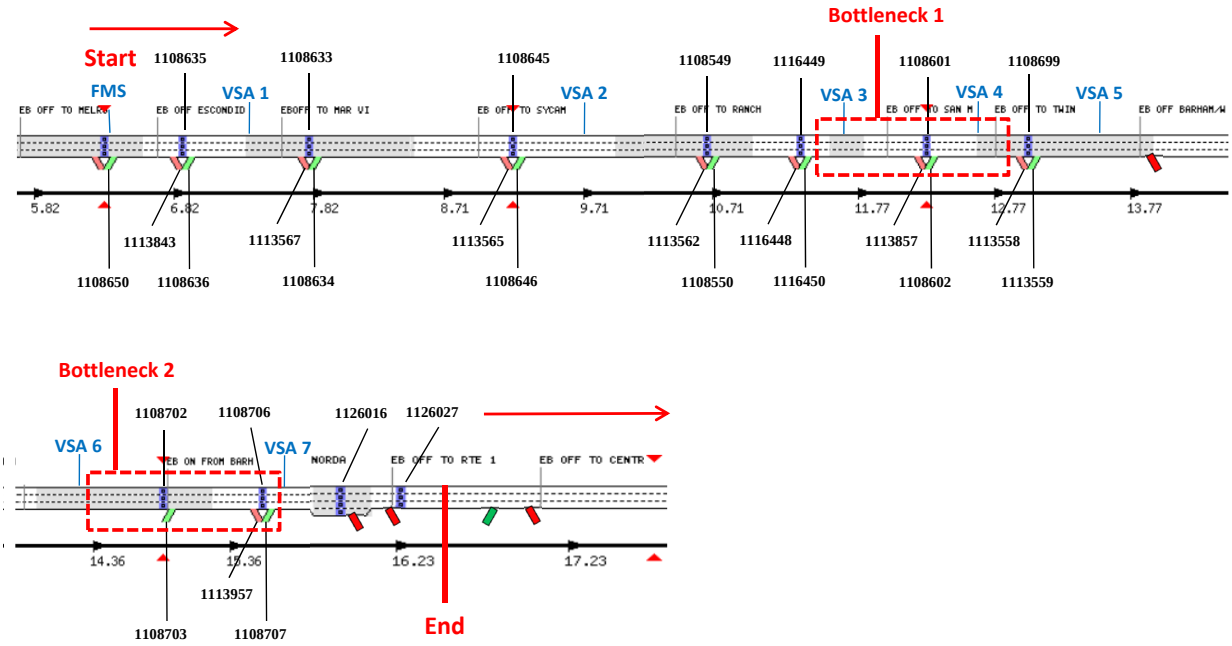


Figure 6.2: SR-78E network configuration and VDS deployment

around PM 11, especially during morning peak hours, since high daily commuters enter the following upstream on-ramps: Sycamore Ave., Las Posas Rd., and San Marcos Blvd. and it causes the onset of bottleneck 1. In summary, based on traffic characteristic studies and bottleneck identifications, the critical location for this VSA test is the most downstream segment around Barham Dr. and Nordahl Rd. Seven VSA signs were placed on the test site. The VSA locations, their posted speed ranges, and the corresponding mainline VDS

number are listed in Table 6.1.

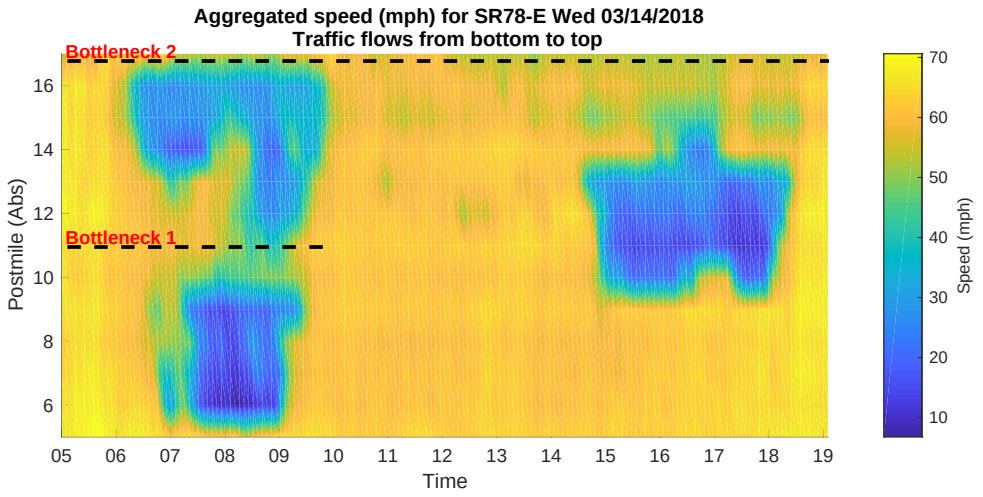


Figure 6.3: SR-78E speed contour, March 14, 2018

Index	Abs. PM	Min. Speed	Max. Speed	VDS	Location	City
FMS 1	6.316	N/A	N/A	1108635	Vista Village Dr.	Vista
VSA 1	6.822	5	65	1108633	Sunset/Escondido Dr.	Vista
VSA 2	9.214	5	65	1108645	Sycamore Ave.	Vista
VSA 3	11.36	5	65	1116449	Las Posas Rd.	San Marcos
VSA 4	12.27	5	65	1108601	San Marcos Blvd.	San Marcos
VSA 5	13.018	5	55	1108699	Twin Oaks Valley Rd.	San Marcos
VSA 6	14.856	5	55	1108702	Barham Dr.	San Marcos
VSA 7	15.593	5	65	1108706	Nordahl Rd.	San Marcos

Table 6.1: VSA locations and its minimum and maximum advisory speed

6.3 Variable Speed Advisory Control Design

Traffic Flow Stabilization

Consider the standard first order model for traffic flow on a single lane freeway parameterized by $x \in [0, L]$ and times t. The dynamics of the density on the freeway is given by Lighthill-Whitham-Richards (LWR) partial differential equation (PDE):

$$\frac{\partial}{\partial t}\rho(x,t) = -\frac{\partial}{\partial x}\big\{\rho(x,t)v(x,t)\big\}. \tag{6.1}$$

The control objective of variable speed limit control is to stabilize traffic flow by commanding a speed profile $v(x,t)$ such that the density profile $\rho(x,t)$ is close to a desired density profile $\rho_d(x,t)$ and the traffic moves at a desired speed $v_d(x,t) > 0$. The desired flow rate is determined by $\phi_d(x,t) = \rho_d(x,t)v_d(x,t)$. For a single lane freeway, the following assumption is introduced.

Assumption 6.3.1.

1. *The speed profile $v(x,t)$ can be commanded.*
2. *The dynamics of the desired density and speed profile satisfies*

$$\frac{\partial}{\partial t}\rho_d(x,t) = -\frac{\partial}{\partial x}\{\rho_d(x,t)v_d(x,t)\}. \tag{6.2}$$

Define the density error $\tilde{\rho}(x,t) = \rho(x,t) - \rho_d(x,t)$ and consider the control law

$$v(x,t) = v_d(x,t) + v_f(x,t) \tag{6.3}$$

$$v_f(x,t) = -\xi(x,t)\frac{\partial}{\partial x}\{v_d(x,t)\tilde{\rho}(x,t)\} \tag{6.4}$$

where $\xi(x,t) \geq 0$.

The speed control laws in equations (6.3) - (6.4) was originally introduced by Li *et al.* in 1997 [57] for stabilizing traffic flow in Automated Highway Systems (AHS). The dynamics for the density error, obtained after substituting the control law in equation (6.3) and (6.4) into (6.1) and using equation (6.2), are

$$\frac{\partial}{\partial t}\tilde{\rho}(x,t) = -\frac{\partial}{\partial x}\{\tilde{\rho}(x,t)v_d(x,t)\} - \frac{\partial}{\partial x}\{\rho(x,t)v_f(x,t)\}. \tag{6.5}$$

For $u\colon [0,L] \to \mathbb{R}$ a real valued function on $[0,L]$, denote the $\mathcal{L}_2$ norm by $||u||_2^2 = \int_0^L u^2 dx$. The following theorem states that with equation (6.3) and (6.4) as the speed control, the desired traffic condition is stable in the $\mathcal{L}_2$ sense.

Theorem 6.3.1. *[57] Consider the single lane freeway model in (6.1). Suppose that the inlet flow rate is $\phi(0,t) = \rho_d(0,t)v_d(0)$, then, under assumption 6.3.1, the control law in (6.3) and (6.4) with $\xi(x,t) \geq 0$ and $\xi(0,t) = \xi(L,t) = 0$, is such that the density error $\tilde{\rho}(x,t) = 0\ \forall x \in [0,L]$ is $\mathcal{L}_2$ stable in time.*

Proof. Consider the following Lyapunov functional:

$$W(t) = \frac{1}{2}\int_0^L \tilde{\rho}(x,t)^2 v_d(x,t)dx. \tag{6.6}$$

Differentiating (6.6) with respect to time and using (6.5),

$$\dot{W}(t) = \int_0^L \tilde{\rho}(x,t)\dot{\tilde{\rho}}(x,t)v_d(x,t)dx = -\int_0^L \tilde{\rho}(x,t)v_d(x,t)\frac{\partial}{\partial x}\{\tilde{\rho}(x,t)v_d(x,t)\} + \tilde{\rho}(x,t)v_d(x,t)\frac{\partial}{\partial x}\{\rho(x,t)v_f(x,t)\}dx. \tag{6.7}$$

The first term in (6.7) is an exact differential:

$$\frac{1}{2}\frac{\partial}{\partial x}\{\tilde{\rho}(x,t)v_d(x,t)\}^2 = \tilde{\rho}(x,t)v_d(x,t)\frac{\partial}{\partial x}\{\tilde{\rho}(x,t)v_d(x,t)\} \tag{6.8}$$

Using the Leibniz rule in the second term of (6.7) and substituting $v_f(x,t)$, this term becomes

$$-\int_0^L \tilde{\rho}(x,t)v_d(x,t)\frac{\partial}{\partial x}\{\rho(x,t)v_f(x,t)\}dx = -\frac{1}{2}\xi(x,t)\rho(x,t)\frac{\partial}{\partial x}\{\tilde{\rho}(x,t)v_d(x,t)\}^2\Big|_0^L - \int_0^L \xi(x,t)\rho(x,t)\Big\{\frac{\partial}{\partial x}\{\tilde{\rho}(x,t)v_d(x,t)\}\Big\}^2 dx. \tag{6.9}$$

Substituting (6.8) and (6.9) into (6.7) yields

$$\dot{W}(t) = -\frac{1}{2}\{\tilde{\rho}(x,t)v_d(x,t)\}^2\Big|_0^L - \frac{1}{2}\xi(x,t)\rho(x,t)\frac{\partial}{\partial x}\{\tilde{\rho}(x,t)v_d(x,t)\}^2\Big|_0^L - \int_0^L \xi(x,t)\rho(x,t)\Big\{\frac{\partial}{\partial x}\{\tilde{\rho}(x,t)v_d(x,t)\}\Big\}^2 dx. \tag{6.10}$$

By the theorem's assumptions, the boundary terms in (6.10) can be cancel out: $\xi(0,t) = \xi(L,t) = 0$ and $\phi(0,t) = v_d(0)\rho_d(0,t)$, $\xi(0,t) = 0$ implies $v(0,t) = v_d(0)$, and therefore $\rho(0,t) = \rho_d(0,t)$. Hence,

$$\dot{W}(t) \leq -\int_0^L \xi(x,t)\rho(x,t)\Big\{\frac{\partial}{\partial x}\{\tilde{\rho}(x,t)v_d(x,t)\}\Big\}^2 dx \leq 0, \tag{6.11}$$

since the density $\rho(x,t) \geq 0$. Then $W(t) \leq W(0)$. Defining $\underline{v_d} = \inf_{(x,t)\in\mathcal{H}} v_d(x,t)$, and $\overline{v_d} = \sup_{(x,t)\in\mathcal{H}} v_d(x,t)$,

$$\underline{v_d}||\tilde{\rho}(\cdot,t)||_2^2 \leq W(t) \leq W(0) \leq \overline{V_d}||\tilde{\rho}(\cdot,0)||_2^2.$$

Thus, for all $t \geq 0$, $||\tilde{\rho}(\cdot,t)||_2 \leq \alpha||\tilde{\rho}(\cdot,0)||_2$ for $\alpha = \sqrt{\overline{v_d}/\underline{v_d}}$, and therefore $\mathcal{L}_2$ stability follows. □

Consider the simplified case when both v_d and ρ_d are constant, the control becomes

$$v(x,t) = v_d - \xi(x,t)v_d\frac{\partial}{\partial x}\rho(x,t). \tag{6.12}$$

If the downstream density is higher than upstream, $\frac{\partial}{\partial x}\rho(x,t) > 0$, the control law (6.12) decreases the speed, which prevents a pile-up downstream. Thus, the control law can be interpreted as a density homogenizing law.

If a freeway segment is divided into N sections for $i = 0, \cdots, N$, let $x = i\Delta x$, $t = k\Delta t$ and $\Delta x = x_{i+1} - x_i$, the discretization of (6.12) is

$$v(i\Delta x, k\Delta t) = v_d - \xi(i\Delta x, k\Delta t)v_d\frac{\rho((i+1)\Delta x, k\Delta t) - \rho(i\Delta x, k\Delta t)}{\Delta x} \tag{6.13}$$

Omitting Δx and Δt and using simplified notation $y(i\Delta x, k\Delta t) := y_i(k)$, (6.13) can be written as

$$v_i(k) = v_d - \xi_i(k)v_d\frac{\rho_{i+1}(k) - \rho_i(k)}{\Delta x} \tag{6.14}$$

If we set $v_d = v_i(k-1)$, (6.14) becomes an integral type controller

$$\bar{v}_i(k) = v_i(k-1) - L_i(k)(\rho_{i+1}(k) - \rho_i(k)) \tag{6.15}$$

where $\bar{v}_i(k)$ is the calculated VSA of section i at time step k, $v_i(k-1)$ is the applied speed command of section i at time step $k-1$, and $L_i(k) = \xi_i(k)v_i(k-1)/\Delta x$ is the control gain. The speed command of section i, v_i, is able to respond to its downstream density changes. In practice, it is necessary to set upper and lower bounds on the speed commands of section i. Then $v_i(k)$ must satisfy the following constraint

$$V_{\min,i} \le v_i(k) \le V_{\max,i} \tag{6.16}$$

where $V_{\max,i}$ is the maximum speed limit allowed in section i, which is usually set to be the default speed limit, and $V_{\min,i}$ is the lowest speed limit we can apply. However, $\bar{v}_i(k)$ calculated by (6.15) may cause unsafe changes of speed limits. Therefore, it is also necessary to apply constraints (6.17) and (6.18) to the advisory speed, such that the speed command does not change too abruptly in time. The speed difference between two consecutive time steps on the same location i cannot be too large,

$$|v_i(k) - v_i(k-1)| \le C_i \tag{6.17}$$

where $C_i \ge 0$ is a positive constant represents the largest change of speed commands allowed between two consecutive time steps in section i. $C_i = 5 \sim 20$ km/h is a suggested value. The speed difference between two consecutive VSA locations during the same time interval k cannot be too large,

$$|v_i(k) - v_{i+1}(k)| \le D_i \tag{6.18}$$

where $D_i \ge 0$ is a positive constant represents the largest change of speed commands allowed during the same time interval k between section i and $i+1$. $D_i = 5 \sim 20$ km/h is a suggested value.

If the average vehicle length is known, the density can be estimated by occupancy measurements [14] and (6.15) can be written as

$$\bar{v}_i(k) = v_i(k-1) - G_i(k)(o_{i+1}(k) - o_i(k)) \tag{6.19}$$

where $o_i(k)$ is the occupancy of section i at time step k and $G_i(k)$ is the control gain.

In practice, traffic shockwave propagation speeds on freeway need to be considered when developing the VSA system. A constant shockwave speed of about 13 mph (or 5.8 m/s) was calculated in the following reference [63] based on NGSIM data. Since the VSA algorithm is based on downstream traffic detection, an occupancy or density increase indicates that congestion happens downstream. If the VSA algorithm looks ahead at least $2 \sim 3$ sections downstream, and the control update time interval is 30 seconds, then the shockwave could propagate about 175 meters and the VSA upstream should be able to respond the density increase shockwave by reducing the feeding flow (via reducing the advisory speed). Therefore, the shockwave was under detection. However, there could be a time delay depending on sensor (loop detector) density; the higher density in traffic detector, the less time delay.

A Heuristic Bottleneck Flow Maximization Strategy

In the previous section, a mainline traffic flow stabilization controller was synthesized using Lyapunov's direct method. The main goal of advisory speed is bottleneck flow maximization. From the controller structure, the advisory speed of section i is determined by the occupancy $o_i(k)$ at section i and its immediately downstream occupancy $o_{i+1}(k)$. It is possible to propose different types of advisory speed strategy based on the occupancy measurements of several upstream sections to section i.

The advisory speed of section i has determined using the next n downstream occupancy measurements of section i, as follows

$$\bar{v}_i(k) = \frac{\alpha_i}{\omega_i(k)}, \tag{6.20}$$

$$\omega_i(k) = p_{i0}o_i(k) + p_{i1}o_{i+1}(k) + \cdots + p_{in}o_{i+n}(k), 0 \le n \le N \tag{6.21}$$

$$p_{i0} + p_{i1} + \cdots + p_{in} = 1 \tag{6.22}$$

where α_i is a parameter and ω_i is the weighted occupancy by parameters $p_{ij}, j = 0, \ldots, n$. The parameter p_{ij} can be determined by

$$p_{ij}(k) = \frac{o_j(k)}{\sum_{k=0}^{n} o_k(k)}, j = 0, \ldots, n. \tag{6.23}$$

Similarly, the advisory speed of section i can be determined based on the speed measurements of the next n downstream sections to section i, as follows

$$\bar{v}_i(k) = \beta_i V_i^w(k), \tag{6.24}$$

$$V_i^w(k) = p_{i0} v_i^m(k) + p_{i1} v_{i+1}^m(k) + \cdots + p_{in} v_{i+n}^m(k), 0 \leq n \leq N \tag{6.25}$$

$$p_{i0} + p_{i1} + \cdots + p_{in} = 1 \tag{6.26}$$

where β_i is a parameter and V_i^w is the weighted speed by parameters $p_{ij}, j = 0, \ldots, n$.

Comparing with the one station look ahead VSA strategy in (6.19), the weighted occupancy strategy in (6.20) or weighted speed strategy in (6.24) can look ahead two or three sections of downstream traffic since they use downstream data in the determination of the current section's VSA. Using multiple downstream data reflects the severity of congestion (the intensity of congestion) more accurate, therefore the VSA in upstream can give proper speed regulation. In addition, the detector data should provide the algorithm with sufficient information to determine the advisory speed. Using multiple downstream data to calculate the VSA has better detector's failure tolerance. If one of the downstream detector is malfunction, the algorithm can still calculate the VSA based on its adjacent downstream detectors. Since the proposed VSA strategies are based on real-time traffic measurements (speed or occupancy), these VSA strategies are able to respond to the change of mainline traffic with proper advised speed. Therefore, they can be applied to the freeway with multiple recurrent or non-recurrent bottlenecks.

Simulation Results

Before pursuing a field test implementation, the Aimsun microsimulation model was used as a test platform to conduct performance evaluations. The proposed VSA algorithms was tested using well calibrated Aimsun model. The following system-wide performance indicators provided by Aimsun were used for evaluation of the algorithm implemented in the simulation model: total travel time (TTT), total travel distance (TTD), total delay (TD) (obtained by deducting hypothetical free-flow travel times from simulated travel times), total number of stops (TNOS) (used as a system-wide performance parameter for traffic smoothness). By comparing the proposed VSA algorithm with the no control case, the improvement of an index x is calculated by

$$\Delta x = \frac{x_{\text{VSA}} - x_{\text{No VSA}}}{x_{\text{No VSA}}}. \tag{6.27}$$

In the simulations, traffic data obtained from March 15, 2017 was selected to drive the simulation model. This data contains all the on-ramps demand, off-ramp split ratio and freeway most upstream mainlines flow. The simulation model was executed during the AM peak hours from 6:00 AM to 9:30 AM. Each day was run for 10 replications (random seeds). The 10 replications for each model day were also different. The results were then averaged over the 10 replications.

By changing the driver compliance and using different VSA controllers, we can evaluate their different freeway performance. For practical purposes, a threshold value of occupancy O_i^T at the i-th VSA was monitored. Thus, if the measured occupancy is greater than the set threshold value of occupancy, the VSA is activated, otherwise the VSA keep the measure speed of its location. O_i^T was suggested to be 0.12. The look ahead VSA control in (6.19) with occupancy threshold is

$$\bar{v}_i(k) = \begin{cases} v_i(k-1) - G_i(k)(o_{i+1}(k) - o_i(k)), \text{if } o_i(k) > O_i^T, i = 1, \cdots, N \\ v_i(k), \text{else} \end{cases} \tag{6.28}$$

The simulation results of using look ahead occupancy based VSA control in (6.28) is shown in Table 6.2.

Using weighted occupancy of next two downstream occupancy, (6.20) with occupancy threshold is

$$\bar{v}_i(k) = \begin{cases} \frac{\alpha_i}{\omega_i(k)}, \text{if } o_i(k) > O_i^T, i = 1, \cdots, N \\ v_i(k), \text{else} \end{cases} \tag{6.29}$$

where α_i is a parameter and weighted occupancy is

$$\omega_i = p_{i0} o_i(k) + p_{i1} o_{i+1}(k) + p_{i2} o_{i+2}(k), \tag{6.30}$$

$$p_{i0} + p_{i1} + p_{i2} = 1. \tag{6.31}$$

The parameters used in the simulation were $(p_{i0}, p_{i1}, p_{i2}) = (0.5, 0.3, 0.2)$. Suppose the maximum flow $f_i^{\max}$ of the mainline is 1800 vehicle per hour per lane. We want to regulate flow at 80% of the maximum flow, that is $0.8 \times 1800 = 1440$. Suppose vehicle density can be estimated by a constant k_i^{occ} times occupancy, that is $\rho_i = k_i^{\text{occ}} o_i$ and $k_i^{\text{occ}} = 1.6$. The operation speed can then be set to

$$\bar{v}_i = \frac{0.8 f_i^{\max}}{\rho_i} = \frac{\alpha_i}{o_i} \tag{6.32}$$

Then, α_i is determined by $\alpha_i = 0.8 f_i^{\max} / k_i^{\text{occ}} = 0.8 \times 1800/1.6 = 900$.The simulation results of using weighted occupancy based VSA control in equation (6.29-6.31) is shown in Table 6.3.

There are several observations that can be drawn from Table 6.2 and Table 6.3. First, higher driver compliance makes TTT, TTD, TD and TNOS improvement better. The more drivers follow the advised speed before entering the most downstream bottleneck, the higher performance the system can reach. It indicates that the VSA controller can regulate the mainline traffic if certain percentage of drivers obey the advised speed. Second, the overall performance in Table 6.2 and Table 6.3 are both improving the freeway performance, but their performance don't have significant difference. A possible explanation for this observation is that the test site has only one bottleneck in the most downstream and both of control

strategies uses same types of data, i.e. occupancy. If the severity of congestion in the next downstream of a VSA station is similar to the further downstream of that VSA, then the advisory speed calculated by only using the next downstream occupancy will be similar to the VSA strategy using further downstream data.

Driver compliance	Δ TTT	Δ TTD	Δ TD	Δ TNOS
10 %	-2.53%	0.39%	-9.4%	-1.8%
25 %	-3.3%	0.4%	-8.1%	-1.8%
50 %	-4.84%	0.81%	-11.74%	-3.6%
Average	-3.56%	0.53%	-9.75%	-2.4%

Table 6.2: System-wide performance indicators changes with look ahead occupancy based VSA control (6.28)

Driver compliance	Δ TTT	Δ TTD	Δ TD	Δ TNOS
10 %	-2.45%	0.42%	-8.92%	-1.8%
25 %	-3.21%	0.45%	-9.32%	-1.7%
50 %	-4.92%	0.78%	-12.72%	-3.5%
Average	-3.53%	0.55%	-10.32%	-2.33%

Table 6.3: System-wide performance indicators changes with weighted occupancy based VSA control (6.29-6.31)

6.4 Field Implementation

The field test was conducted from Tuesday, March 20, 2018 to Friday, May 4, 2018. From March 20, 2018 to April 9, 2018, the project team was deploying VSA devices in the freeway and adjusting the system parameters. During this period, the team was fixing bugs, and making necessary modifications and adjustments to the VSA algorithm. In this field test, the advisory speed limit would be varied every 30 seconds in 5 mph increments. Further, the upper limit of the advisory speed limit of each sign was set to 65 mph, while that of lower limit of the advisory speed limit was set to 5 mph. These algorithms were implemented by C code, which determined the advisory speed limit values for all seven signs using the speed measurements uploaded from each loop detector station and each radar detector every 30 seconds through a wireless communication network.

Hardware Architecture

The variable speed advisory (VSA) sign specifically designed for this research in cooperation with the project team in PATH [85] and engineers at Traffic Logix [103]. Traffic Logix 15

inch radar sign SafePace 625 are used as the VSA display device in this field test. These signs are LED panel powered by solar power and backup battery and it also contains a radar speed detector and a communication module that transmit speed data to the website. Figure 6.4 shows one of the VSA sign, during the VSA field test, a total of seven such signs were manufactured and installed along the SR-78E test site. Seven Variable Message Signs (VMS) with modified firmware were distributed along a 10.8-mile stretch of California State Route 78 near San Diego, which is shown im Figure 6.4. The firmware and back-end were modified by the manufacturer (TrafficLogix Corp.) to accommodate the faster update rate needed for timely traffic data acquisition and sign control. The TrafficLogix Safepace 650 Variable Message Sign was used in this project for several reasons:

1. Capability of changing the speed advisory remotely.
2. Integrated radar for sensing traffic speed.
3. Receipt of radar statistics at 30-second update rate.
4. Speed display update rate of 30 seconds.
5. Solar powered charging system that allowed for continuous operation.

In its standard configuration, suitable for normal traffic operations, the Safepace 650 has an update interval of 5 minutes. Simulation studies performed at PATH indicated that traffic patterns can change much faster than 5 minutes, so the engineering staff at TrafficLogix was asked to modify the firmware on the signs and the back-end server to allow for 30 second intervals. They also provided an applications programming interface that allowed software control by the user. When this was done, the signs were successfully deployed.

Software Architectures

Software was constructed as a sequence of processes executed by a parent script. The child processes communicated with each other via a publish/subscribe database, temporary data files, and php scripts. This software structure was largely dictated by two different sources of data (a data feed from Caltrans containing occupancy, flow, and speed over the loops embedded in the highway, and radar data from the signs transmitted from the TrafficLogix data server located in New York) and one control channel (Variable Speed Advisories that were output from the control algorithm and sent to the web server in New York for transmission to the VSA signs).

Variable speed advisories were calculated using an algorithm that applied the most recent data from the two data sources, namely radar speed from the VSA signs and loop data from the Caltrans data feed. The occupancy measurement based VSA rules in equations (6.19)-(6.23) and its constrains in equations (6.16)-(6.18) was implemented in this field test. Since the update rate of the two data streams was 30 seconds, the data was read and analyzed every 30 seconds. Upon calculation of the variable speed advisories, the VSAs were sent to

the TrafficLogix application server for display on the VSA signs. The software architecture is shown in Figure 6.5. VDS data from loop detectors (bottom data stream) was collected by Caltrans into an XML file and sent to PATH. Radar statistics from the VSA signs (upper data stream) were received from TrafficLogix' web server and forwarded to PATH. Variable Speed Advisories are calculated in the PATH server and sent to the VSA signs via the TrafficLogix server. The data collected from the field, as well as the VSAs, were sent to the VSA web page for display.

Figure 6.4: TrafficLogix Safepace 650 Variable Message Sign

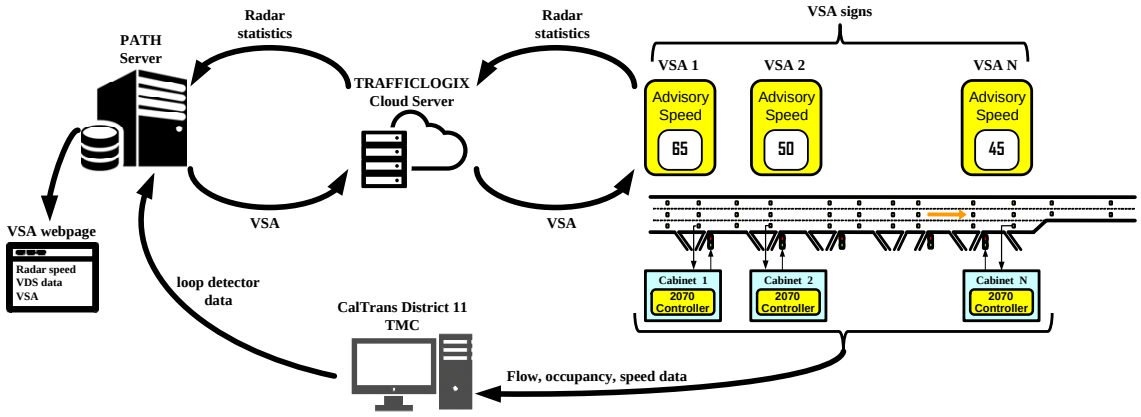


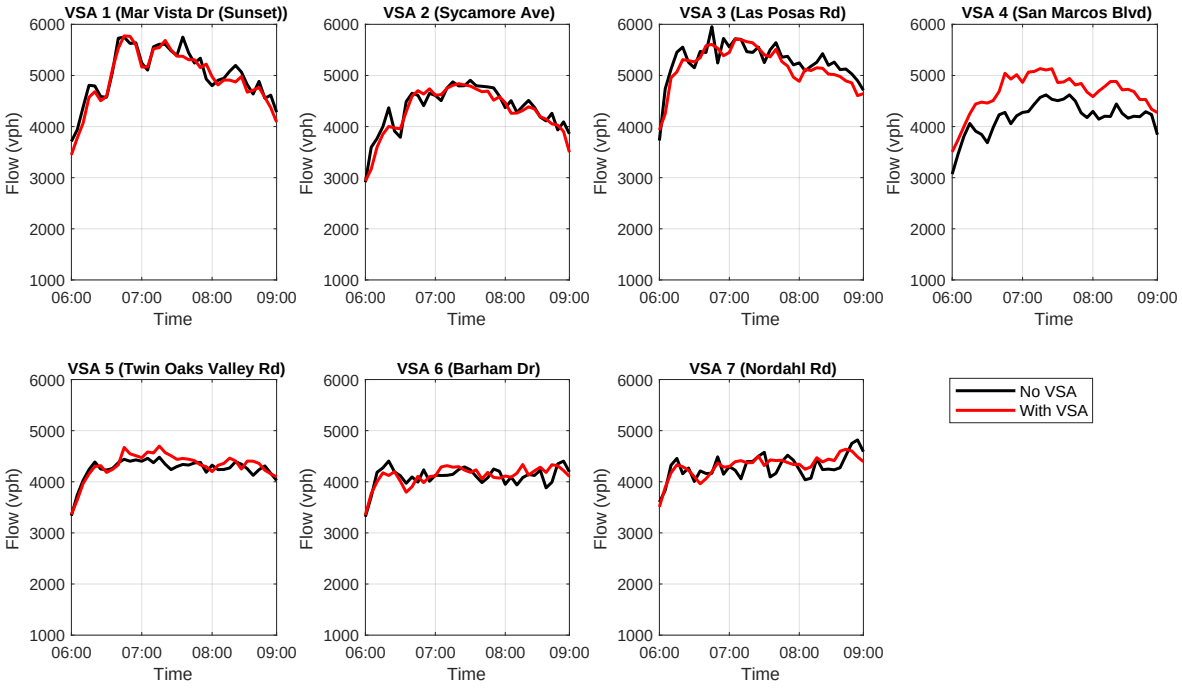
Figure 6.5: Software architecture of the variable speed advisory system

6.5 Field Test Results

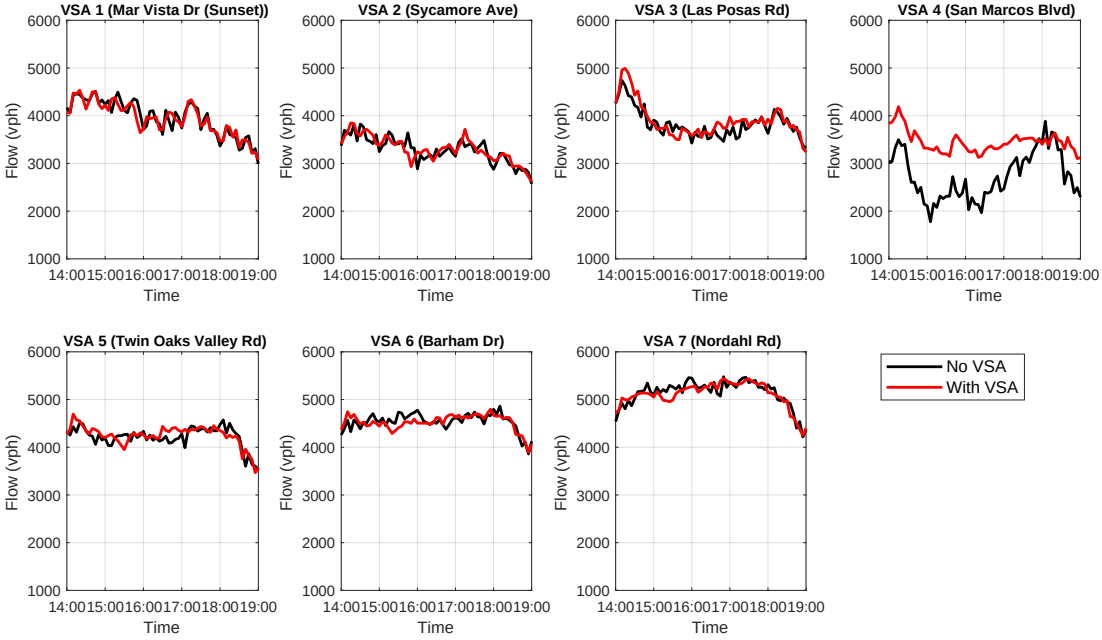
All seven VSA devices were installed at the SR-78 Eastbound test site before March 16, 2018. After a 3-week system tuning period from March 19, 2018 to April 8, 2018, the system was finally activated at approximately 6:00 AM on April 9, 2018, and the data for the "after" period was collected until May 11, 2018.

Flow Study

Since traffic flow fluctuates day to day, this study averages the flow on weekdays during the field test to compare the change in flow patterns under VSA control. For the no-control period, no VSA signs were placed on the freeway shoulder. In the field test period, VSA signs were placed and activated during morning peak hours (from 6:00 AM to 9:00 AM) and evening peak hours (from 2:00 PM to 7:00 PM). The VSA field test were implemented from April 9 to May 4, 2018. Recurrent morning congestion happened at San Marcos Blvd. (the location of VSA 4 near bottleneck 1) and at Nordahl Rd. (the location of VSA 7 near bottleneck 2). Figure 6.6 (a) and (b) show 5 minutes aggregated flow from all mainline loop detectors near VSA during morning and evening peak hours, respectively. These figures present no VSA control (average weekday flow from May 7 to May 11, 2018) and VSA control (average weekday flow from April 9 to May 4, 2018) cases, respectively. Except for the flow near San Marcos Blvd. (VSA 4), these plots indicate that the flow patterns are similar in the no control and VSA control cases, which indicates that traffic demand was stable and comparable in the no control and VSA control cases. As shown in Figure 6.6 (a), the average flow of morning peak hours near San Marcos Blvd. (VSA 4) increased from 4100 vph to 4500 vph after the VSA deployment and the average flow near Nordahl Rd. (VSA 7) also increased slightly. As shown in Figure 6.6 (b), the average flow of evening peak hour near San Marcos Blvd. (VSA 4) increased from 2700 vph to 3200 vph. This observation indicates that the traffic flow near San Marcos Blvd. (the midstream bottleneck 1) was improved.



(a)



(b)

Figure 6.6: Comparison of flow between no control and VSA control: (a) during morning peak (b) during evening peak

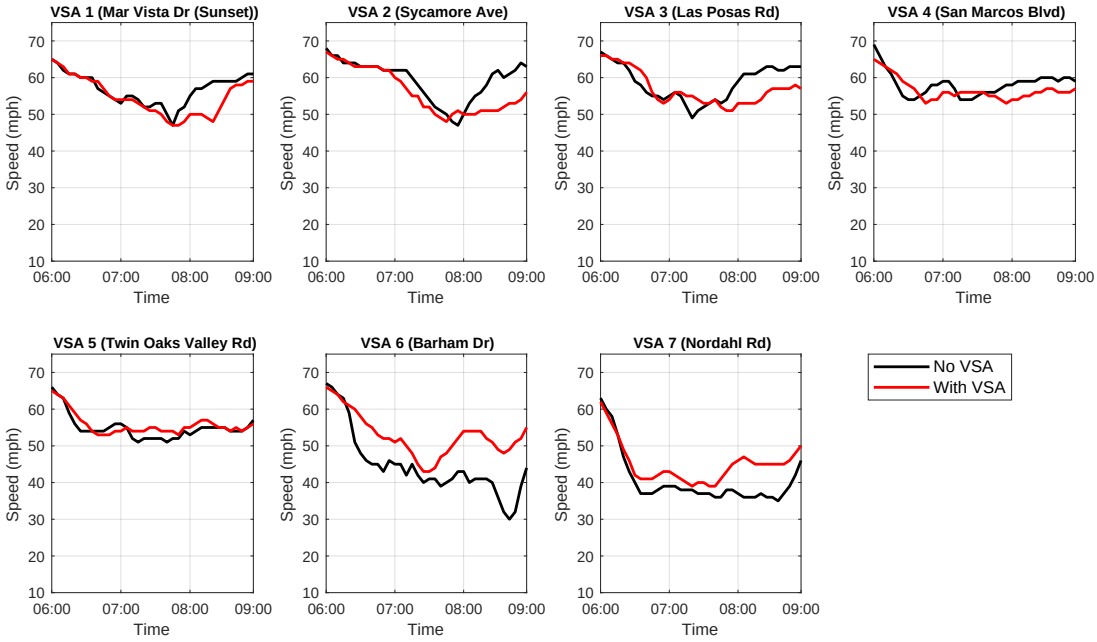
Speed Study

When vehicles pass a congestion bottleneck, the deceleration and acceleration cause a capacity drop. Ideally, if drivers follow the advisory speed posted by VSA, the upstream discharge flow can be reduced by lowering the driving speed limit, and then gradually raising speed limits after vehicles pass downstream bottlenecks. In this way, VSA reduces the occurrence of congestion, stop-and-go conditions, shock waves, and capacity drop. Figure 6.7 (a) and (b) present the 5 minutes aggregated speed profile at each VSA location during morning and evening peak hours, respectively. These figures present no VSA control (average weekday speed from May 7 to May 11, 2018) and VSA control (average weekday speed from April 9 to May 4, 2018) cases, respectively. The morning mainline speed profile in Figure 6.7 (a) shows that downstream speeds at the VSA 5, VSA 6, and VSA 7 locations under VSA control were higher than those under no control scenario. For the upstream mainline speeds at the VSA 2, VSA 3 and VSA 4 locations, although the speeds were lower than those under no control scenario, the upstream speeds were maintained around 55 mph from 7:30 AM to 9:00 AM, which is desirable since the upstream speed slowed down and kept a stable speed to delay the downstream congestion onset. Therefore, the downstream bottleneck (bottleneck 2) the downstream bottleneck speed were increased by VSA control and it also prevented drastic speed drops in the most downstream during the morning peak hours. The evening mainline speed profile in Figure 6.7 (b) shows that speed in all VSA locations under VSA control was only slightly higher than those under no control scenario. However, the speed variations at the VSA 4, VSA 5, VSA 6 locations from 3:00 PM to 6:00 PM under VSA control were lower than those no control scenario, which indicates that advisory speeds can smooth speed transitions.

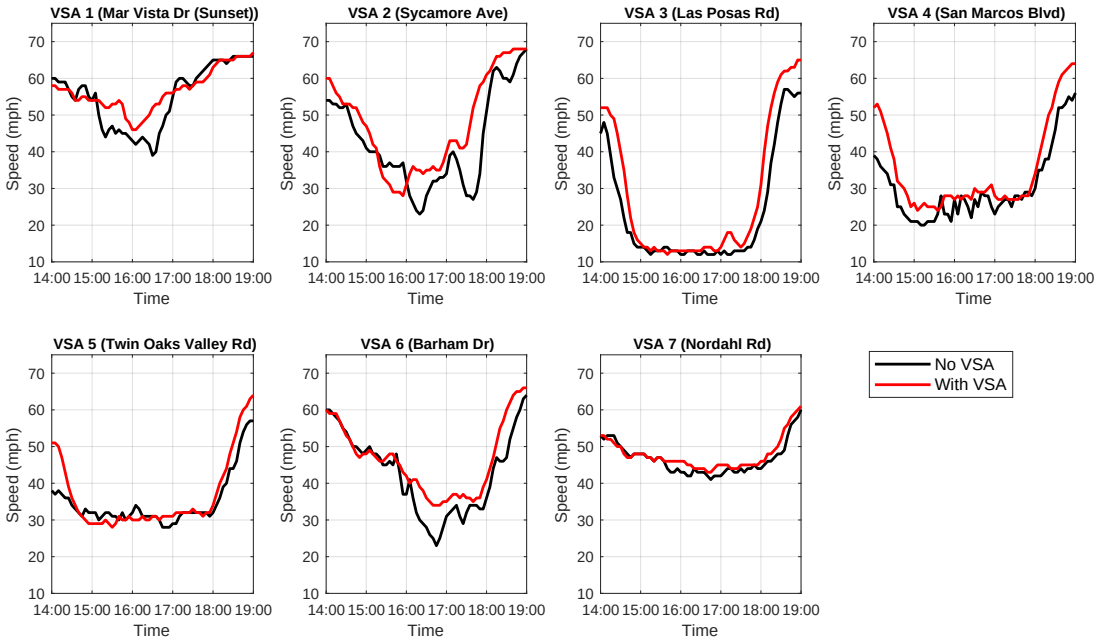
Figure 6.8 (a) and (b) present the 30 second speed data and advisory speed at each VSA location on April 26, 2018 during morning and evening peaks, respectively. The morning peak hours from 7:00 AM to 9:00 AM in Figure 6.8 (a) show that the advisory speed at the VSA 7 location became lower than the measured speed when the most downstream congestion onset at 6:30 AM and those upstream VSA also recommended deceleration from 6:30 AM to 8:00 AM. The speeds during the evening peak in Figure 6.8 (b) shows that the advisory speed was generally lower than the measured speed when congestion onset at the VSA 4, VSA 5, VSA 6, and VSA 7 locations from 3:00 PM to 5:30 PM and then the advisory speed was gradually raising after 5:30 PM (the congestion dissipates). The comparison in Figure 6.8 shows that the advisory speed is reasonable and achievable, which encourage drivers to follow the VSA and therefore improves freeway mobility, rather than confuse them and reduce traffic performance.

Compliance per VSA speed

Figure 6.9 (a) and (b) shows a box plot of the speed compliance at each of the advisory speeds from April 23, 2018 to April 27, 2018 for both morning and evening peak hours,

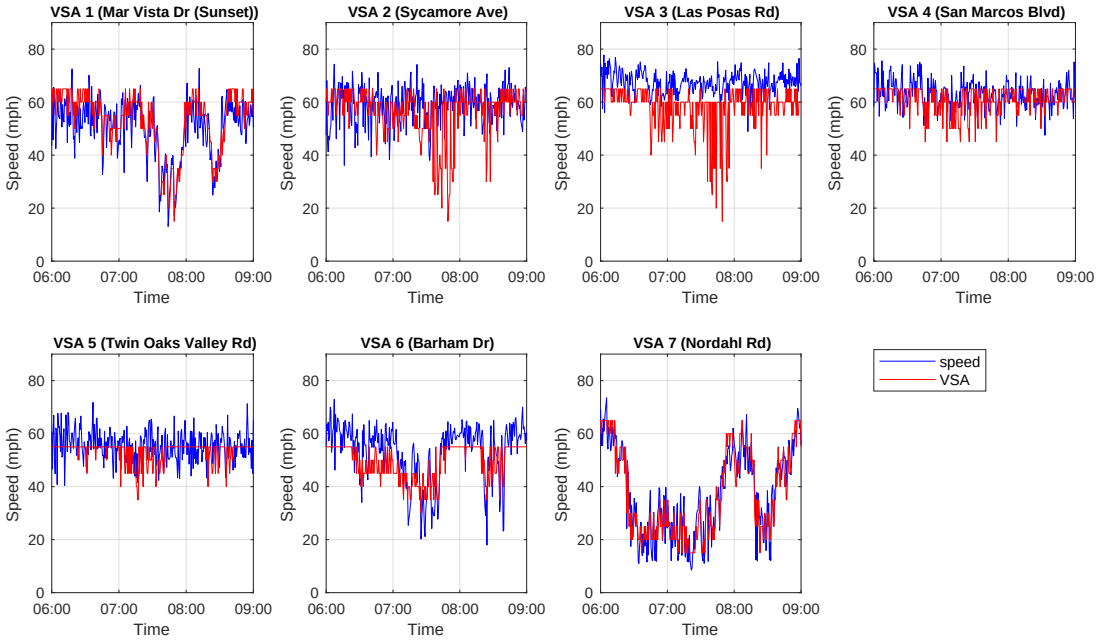


(a)

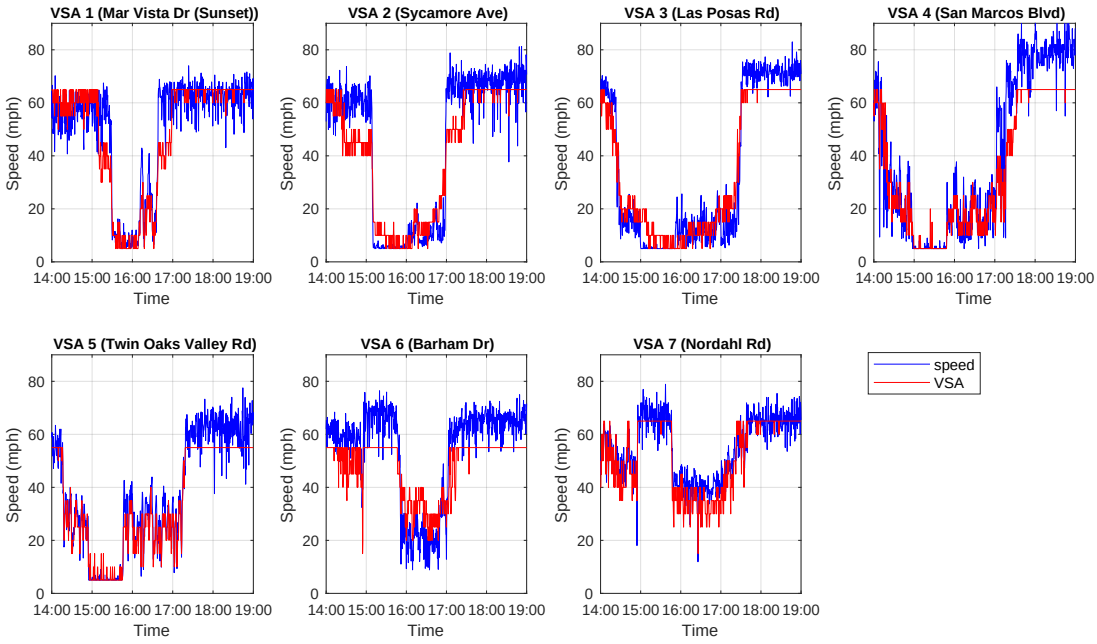


(b)

Figure 6.7: Comparison of speed between no control and VSA control: (a) during morning peak (b) during evening peak



(a)



(b)

Figure 6.8: Comparison between speed profiles and advisory speed (30 second data) in April 26, 2018: (a) during morning peak (b) during evening peak

respectively. The driver compliance $s_i(k)$ of section i at time step k is defined as

$$s_i(k) = \bar{v}_i(k) - v_i^m(k) \tag{6.33}$$

where $\bar{v}_i(k)$ is the VSA of section i and $v_i^m(k)$ is measured speed of section i. The median compliance slightly less than zero means that vehicles are traveling at speeds lower than the advisory speed. At times when the VSA displays a speed above 50 mph (65, 60, 55, and 50 mph), the compliance level is between 10 mph and -10 mph, which indicates that the posted speed for drivers is feasible.

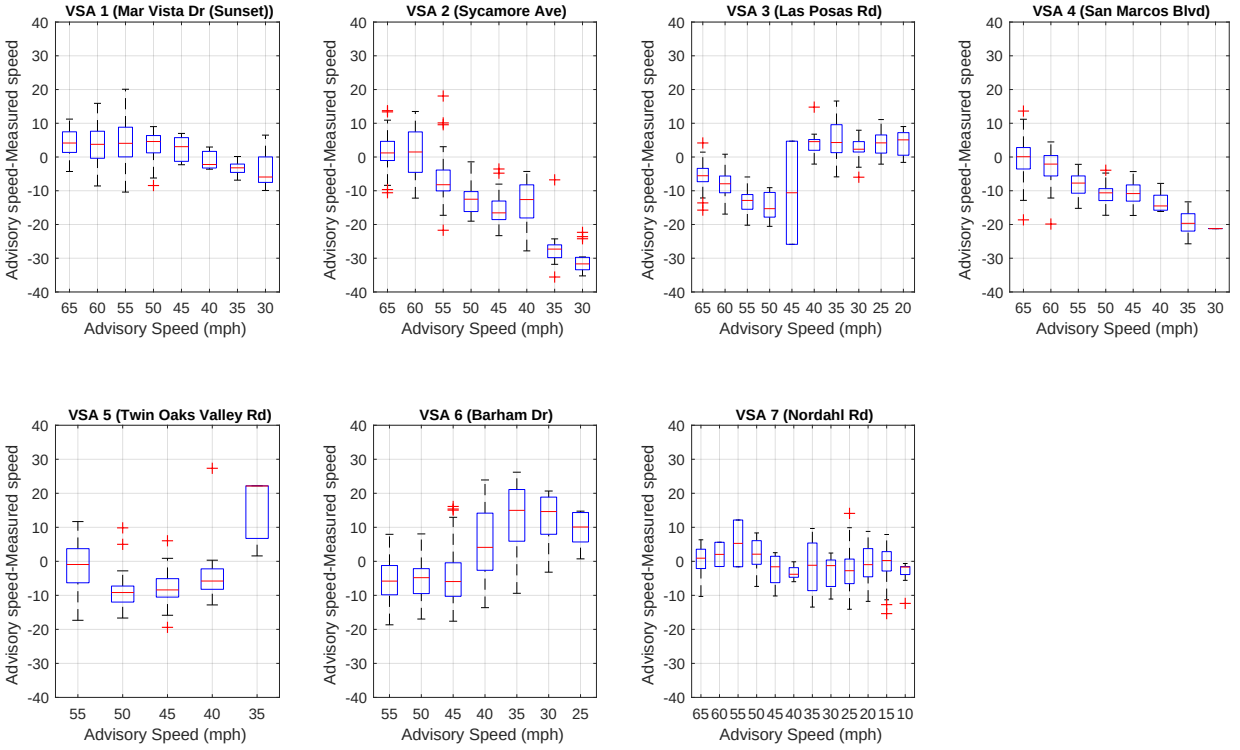
Variable Speed Advisory Profile

The variable speed advisory calculated by the algorithm varied by both freeway segment and the traffic conditions during the field test period. Figure 6.10 (a) and (b) shows the maximum advisory speed given for any of the seven freeway segment zones was 65 mph except for locations 5 and 6 has maximum advisory speed in 55 mph and the minimum of advisory speed is 5 mph, which indicate the message sign post the expected maximum and minimum values. Comparing the average measured speed (blue line with stars) with posted average VSA speed (red line with circles) in Figure 6.10 (a) and (b), it shows that average speed is close to the posted VSA in average, which indicates that most drivers in the test site can follow the posted speed smoothly.

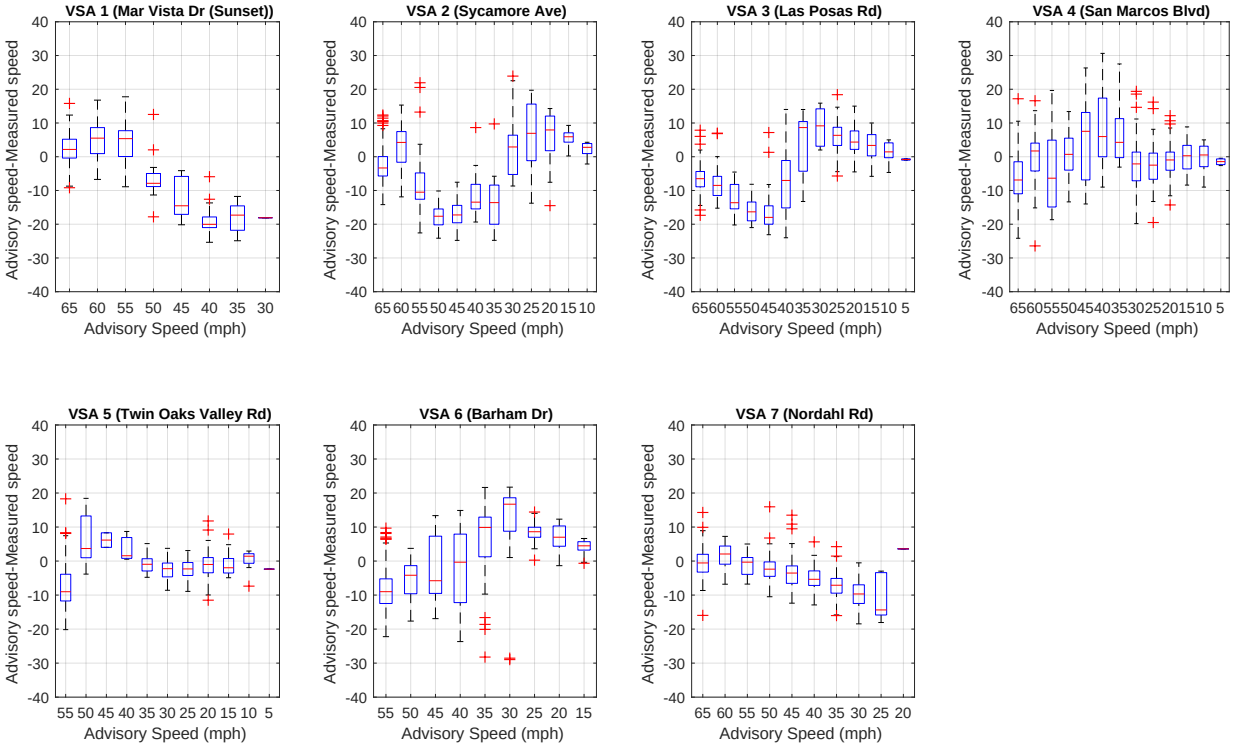
Figure 6.11 (a) and (b) shows the speed difference between two consecutive VSA signs during morning and evening peak, respectively, which is calculated by

$$\Delta v = v_{i+1} - v_i, i = 1, \cdots, 6 \tag{6.34}$$

where v_i is the advisory speed of the i-th VSA and v_{i+1} is the advisory speed of the downstream of the i-th VSA. Figure 6.11 shows that the changes between the advisory speed are around 10 mph during 80 percent of the test time, which indicates that drivers are requested a speed reduction less than 10 mph and therefore the change of advisory speed is acceptable.

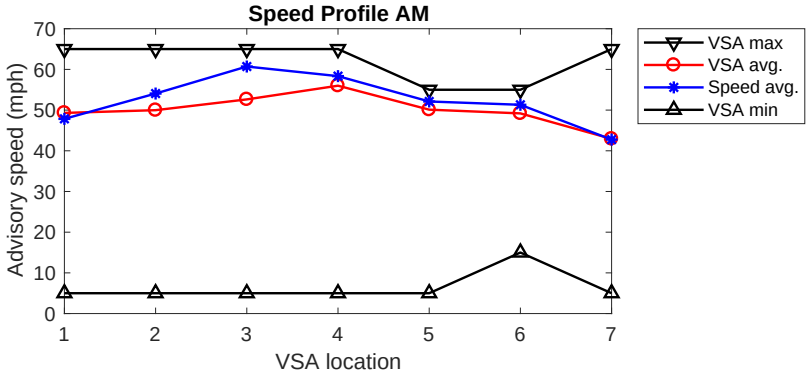


(a)

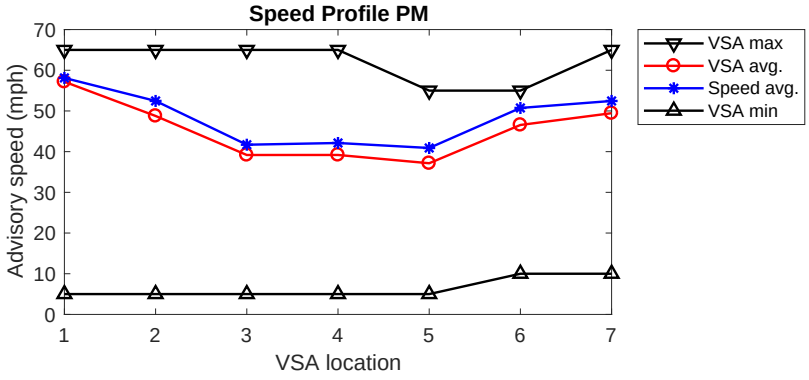


(b)

Figure 6.9: Compliance (advisory speed minus measured speed) in 30 second data from 4/23/2018 to 4/27/2018: (a) during morning peak (b) during evening peak



(a)



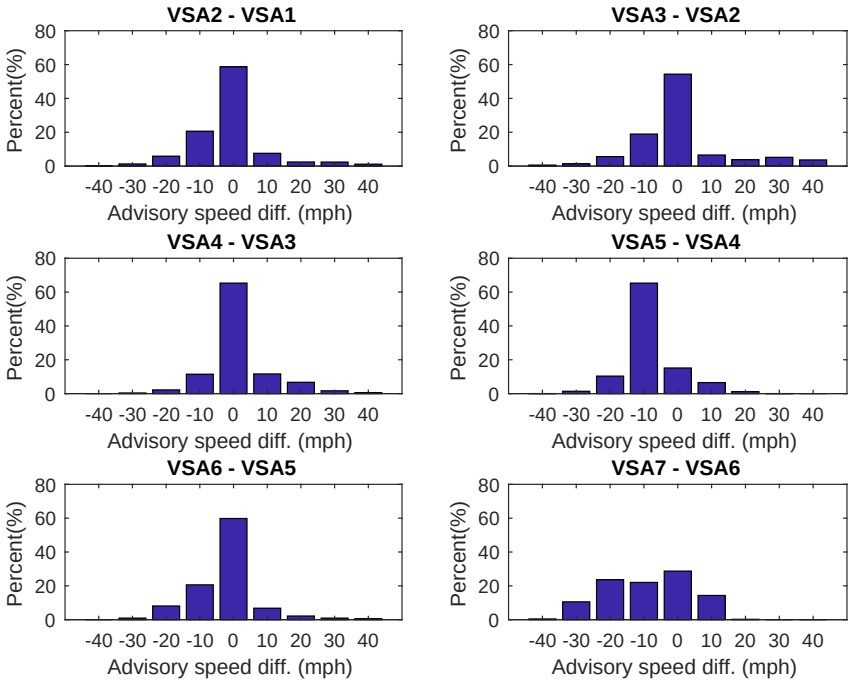
(b)

Figure 6.10: VSA profile: (a) during morning peak (b) during evening peak

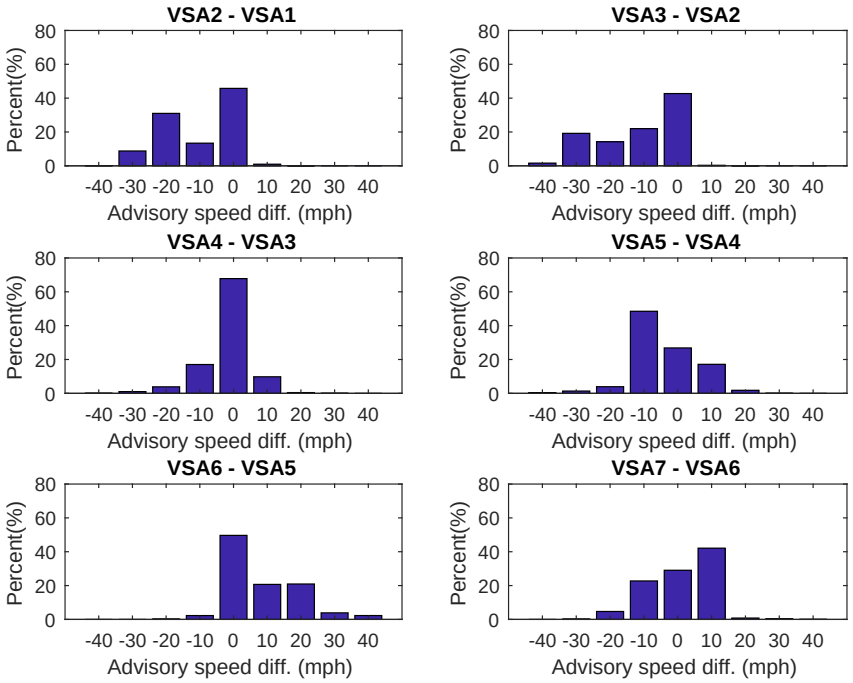
Freeway Performance Study

Variable advisory speed (VSA) is introduced mainly for alleviating freeway congestion and improving safety by homogenizing vehicle speeds. In order to study the influence of VSA control, three main freeway performance indices: VMT (5.15), VHT (5.16), and Q (5.17) are considered in this research. The data for freeway performance evaluation is obtained from PeMS [86]. The data for the segment of test site SR-78E from postmile 6.316 to postmile 15.593 are used. Hourly data VMT, VHT, and Q are used in this study. The VSA test duration is the morning peak hours and afternoon peak hours in the weekdays from April 9, 2018 to May 4, 2018, four weeks of VSA test in total. The AM VSA activation time is from 6:00 AM to 9:00 AM. The PM VSA activation time is from 2:00 PM to 7:00 PM. Two weeks of weekday data (without VSA control): from March 12, 2018 to March 16, 2018 and from May 7, 2018 to May 11, 2018 are selected as baseline for freeway performance comparison. The improvement of an index x is computed by

$$\Delta x = \frac{x_{\text{with VSA}} - x_{\text{without VSA}}}{x_{\text{without VSA}}}. \tag{6.35}$$



(a)



(b)

Figure 6.11: Advisory speed changes between adjacent signs: (a) during morning peak (b) during evening peak

Table 6.4, Table 6.5, and Table 6.6 show average VMT, VHT, and Q along the test site during morning and evening field test period, respectively. The morning average VMT is 50957.70 Veh-Miles and the evening average VMT is 47261.71 Veh-Miles. The morning average VHT is 979.07 Veh-Hours and the evening average VHT is 1208.45 Veh-Hours. The morning average Q is 53.22 mph and the evening average Q is 44.65 mph.

Test week, VMT (Veh-Miles)	6-7 AM	7-8 AM	8-9 AM	2-3 PM	3-4 PM	4-5 PM	5-6 PM	6-7 PM
4/9/2018 - 4/13/2018	54429.26	50702.52	46996.4	48852.64	50468.84	51000	47226.16	40287.66
4/16/2018 - 4/20/2018	55351.58	50822.68	47635.8	50646.44	50315.38	50742.28	45592.64	37796.38
4/23/2018 - 4/27/2018	55419.54	51389.82	46597.72	48593.16	47882.56	50223.76	46286.16	37519.84
4/30/2018 - 5/4/2018	54871.28	51273.84	46002.02	52655.9	52344.86	52429.5	47111.34	37258.7
Average VMT	55017.91	51047.21	46807.98	50187.03	50252.91	51098.88	46554.07	38215.64

Table 6.4: Summary of both AM and PM hourly VMT during field test

Test week, VHT (Veh-Hours)	6-7 AM	7-8 AM	8-9 AM	2-3 PM	3-4 PM	4-5 PM	5-6 PM	6-7 PM
4/9/2018 - 4/13/2018	1129.22	966.26	806.22	1710.02	1687.46	1554.5	915.9	626.96
4/16/2018 - 4/20/2018	1137.18	1004	816.7	1554.98	1630.1	1434.7	819.48	575.58
4/23/2018 - 4/27/2018	1096.34	1031.12	776.12	1482.34	1712.56	1538.96	877.26	569.8
4/30/2018 - 5/4/2018	1152.66	1041.02	792.06	1482.22	1417.74	1219.1	799.04	560.38
Average VHT	1128.85	1010.6	797.775	1557.39	1611.965	1436.815	852.92	583.18

Table 6.5: Summary of both AM and PM hourly VHT during field test

Test week, Q (mph)	6-7 AM	7-8 AM	8-9 AM	2-3 PM	3-4 PM	4-5 PM	5-6 PM	6-7 PM
4/9/2018 - 4/13/2018	48.71	53.21	58.45	29.44	30.79	33.62	52.29	64.28
4/16/2018 - 4/20/2018	49.76	51.58	58.44	33.52	31.46	35.86	55.81	65.73
4/23/2018 - 4/27/2018	51.13	51.63	60.15	32.82	29.67	34.07	53.69	65.84
4/30/2018 - 5/4/2018	47.73	49.68	58.22	36.38	38.38	43.79	59.08	66.58
Average Q	49.33	51.52	58.81	33.04	32.57	36.83	55.21	65.60

Table 6.6: Summary of both AM and PM hourly Q during field test

Baseline 3/12/2018 - 3/16/2018	6-7 AM	7-8 AM	8-9 AM	2-3 PM	3-4 PM	4-5 PM	5-6 PM	6-7 PM
VMT (Veh-Miles)	51713.88	48239.8	46757.66	52521.58	50946.14	51431.44	46941.72	37900.66
VHT (Veh-Hours)	1340.3	1267.5	869.24	1470.82	1507.84	1453.76	868.18	642.64
Q (mph)	39.8	42.3	54.77	36.29	33.86	35.47	54.23	60.86

Table 6.7: Summary of both AM and PM hourly VHT, VMT and Q during March 12, 2018 - March 16, 2018

In order to study the field test result, the data without VSA activation before the VSA field test, listed in Table 6.7 and the data without VSA activation after the VSA field

Baseline 5/7/2018 - 5/11/2018	6-7 AM	7-8 AM	8-9 AM	2-3 PM	3-4 PM	4-5 PM	5-6 PM	6-7 PM
VMT (Veh-Miles)	54502.2	51302.9	45197.48	50323.46	48961.12	50683.56	46360.28	37512.9
VHT (Veh-Hours)	1134.48	1015.3	741.84	1520.7	1639.06	1534.4	946.18	582.56
Q (mph)	48.50	51.02	60.97	33.77	30.33	33.32	50.07	64.76

Table 6.8: Summary of both AM and PM hourly VHT, VMT and Q during May 7, 2018 - May 11, 2018

test, listed in Table 6.8 are selected as baselines for evaluation. The comparison of freeway performance between the average of four VSA field test week and baseline data is contained in Table 6.9 (using data from March 12, 2018 to March 16, 2018 as a baseline) and Table 6.10 (using data from May 7, 2018 to May 11, 2018 as a baseline), where **bold text** indicates improvement of the networks performance and *italic text* means deterioration of performance. In Table 6.9, the morning average ΔVMT is 4.10% and the evening average ΔVMT is -1.28%; the morning average ΔVHT is -14.75% and the evening average ΔVHT is 0.12%; the morning average ΔQ is 17.71% and the evening average ΔQ is 0.14%. In Table 6.10, the morning average ΔVMT is 1.33% and the evening average ΔVMT is 1.09%; the morning average ΔVHT is 2.19% and the evening average ΔVHT is -3.07%; the morning average ΔQ is -0.27% and the evening average ΔQ is 5.47%. By averaging the improvement of performance index in Table 6.9 and Table 6.10, the overall performance evaluation results are summarized as follows:

Morning variable speed advisory performance:

- VMT increased by 2.72% on average.
- VHT decreased by 6.28% on average.
- Q increased by 8.71% on average.

Evening variable speed advisory performance:

- VMT decreased by 0.096% on average.
- VHT decreased by 1.47% on average.
- Q increased by 2.80% on average.

The results of the performance analysis illustrated an improvement in all three PeMS performance measures. During the AM peak hours (from 6:00 AM to 9:00 AM), VMT increased by 2.72%, VHT decreased by 6.28%, and Q increased by 8.71%. In PM peak hours (from 2:00 PM to 7:00 PM), two of the three performance measures improved. VMT did not have noticeable improvement, while VHT decreased by 1.47% on average, and therefore Q increased by 2.80% on average. On SR-78E the PM peak hours experiences higher traffic demand, as well as a greater percentage of non-commuters. These two factors may have had some influence in the lower PM performance improvements. As for driver compliance, it

gradually improved as the test progressed and the increase in driver compliance was generally in line with an improvement in system performance.

	6-7 AM	7-8 AM	8-9 AM	2-3 PM	3-4 PM	4-5 PM	5-6 PM	6-7 PM
Δ VMT	**6.38**%	**5.81**%	**0.10**%	*−4.44*%	*−1.36*%	*−0.64*%	*−0.82*%	**0.83**%
Δ VHT	**−15.77**%	**−20.26**%	**−8.22**%	*5.88*%	*6.90*%	**−1.16**%	**−1.75**%	**−9.25**%
Δ Q	**23.95**%	**21.80**%	**7.38**%	*−8.95*%	*−3.79*%	**3.84**%	**1.82**%	**7.80**%

Table 6.9: Summary of both AM and PM hourly performance comparison, comparing to March 12, 2018 to March 16, 2018

	6-7 AM	7-8 AM	8-9 AM	2-3 PM	3-4 PM	4-5 PM	5-6 PM	6-7 PM
Δ VMT	**0.94**%	*−0.49*%	**3.56**%	*−0.27*%	**2.63**%	**0.81**%	**0.41**%	**1.87**%
Δ VHT	**−0.49**%	**−0.46**%	*7.54*%	*2.41*%	**−1.65**%	**−6.35**%	**−9.85**%	*0.10*%
Δ Q	**1.71**%	**0.98**%	*−3.53*%	*−2.16*%	**7.40**%	**10.54**%	**10.28**%	**1.30**%

Table 6.10: Summary of both AM and PM hourly performance comparison, comparing to May 7, 2018 to May 11, 2018

6.6 Summary

In this study, the variable speed advisory freeway traffic management based on occupancy and speed measurement was developed and implemented at the stretch of SR-78E between Vista Village Drive and the interchange with US-15, which is one of the most congestion interchanges in San Diego. At the test site, seven VSA signs, which were updated every 30 seconds in real-time, were deployed at the shoulder of the road for advisory speed display. The VSA system was activated during the morning peak periods of 6:00 AM - 9:00 AM and evening peak periods of 2:00 PM - 7:00 PM on weekdays. After the control system was implemented, a progressive test procedure was conducted for the field tests. First, a dry-run (VSA was calculated and saved but not displayed) was conducted for the first 2 weeks to verify that the overall system was running well. Then, after 4 weeks of tests and data collection, PeMS hourly VHT and VMT data were used for performance evaluation. It was noted that PeMS data is completely independent from the data in PATH computer obtained directly from the 2070 controllers in the field. By using PeMS data, the project team intended to obtain objective performance results to the greatest extent possible. To address demand fluctuation, the freeway efficiency (VMT to VHT ratio) was used as the performance parameter, which could be understood as the average speed. We believed that this ratio could reasonably accommodate traffic demand fluctuation. The data used for VSA OFF was collected in the week (3/12-3/16, 2018) right before VSA signs were mounted, and the week (5/7-5/11, 2018) right after VSA signs were removed. The aggregated data over four

weeks when VSA was ON compared to the data with VSA OFF for average speed (freeway efficiency) increased by 8.71% for morning peak hours, which indicates an improvement in congested traffic. For evening peak hours, average speed (freeway efficiency) increased by 2.8%, which is marginal. This suggests that the algorithm is effective for congested traffic caused by high demand. The simplicity of the advisory speed strategy developed in this study and the flexibility of the hardware/software system used for the field test indicate the possibility of adopting the proposed variable advisory speed system as one of the regular tools for freeway management.

Chapter 7

Signalized Intersections Offset Optimization

7.1 Introduction

In arterial road networks, vehicle idling time at signalized intersections is a major source of delays and emissions. Minimization of idling time can be achieved by aligning signal offsets to encourage sequences of green lights along heavily utilized roads. The offset optimization problem as proposed in [21] approximated queueing processes with sinusoids and sought to minimize queue oscillations which appear with ill times offsets. The problem is amenable to semidefinite relaxation (SDP), which was used to both certify a lower bound on the minimization problem and as a heuristic to obtain near optimal offsets.

For larger networks with thousands of intersections, interior point solvers used in SDPs typically run out of memory. The Burer-Monteiro (BM) method is suggested for solving SDPs to avoid the expensive conic constraints and dramatically reduce the number of optimization variables [11]. This method is a non-linear heuristic for solving SDPs with low rank solutions with surprisingly good results in practice; prior uses of the BM method include e.g. matrix completion [89] and angular synchronization [97].

This chapter presents two examples of demonstrating both the Burer-Monteiro algorithm's scalability as well as improvements to common traffic metrics in a microscopic simulation. The first synthetic example with 1771 intersections in the Manhattan borough of New York City, highlights the scalability of the Burer-Monteiro algorithm to a network with thousands of intersections. Out second example with 420 intersections is a microscopic simulation of Pasadena, a suburb of Los Angeles, California. This simulation uses real green splits, cycle times, and empirically observed demand and was provided by the Connected Corridors project at UC Berkeley [20]. The optimized offsets outperform existing offsets, despite modeling assumptions such as a universal cycle time, sinusoidal queueing dynamics, and constant travel time. Specifically the optimized offsets reduce the delay and queue occupancy.

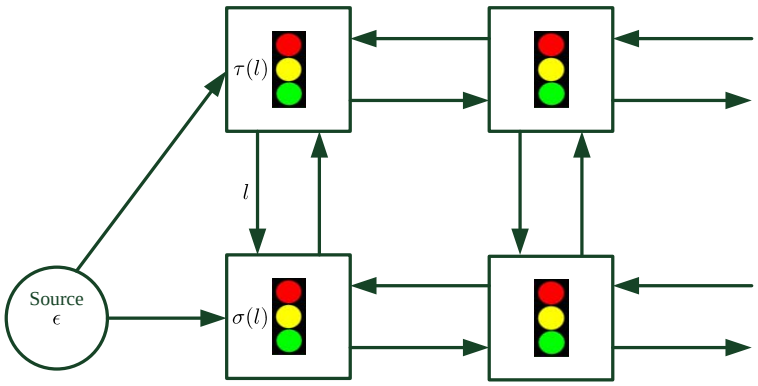


Figure 7.1: Example network

7.2 Traffic Flow Model of Intersections

We describe the routing model and queue dynamics adopted previously for the offset optimization problem by Coogan et al. [21].

A traffic network is modeled as a directed graph $G = (\mathcal{S} \cup \{\epsilon\}, \mathcal{L})$ with a set of links/roads $\mathcal{L}$ and nodes/signalized intersections $\mathcal{S}$. Each link has a vehicle queue where vehicles wait to be discharged. Let $\sigma : \mathcal{L} \mapsto \mathcal{S}$ map a link to the signalized intersection that actuates a link's queue. Similarly, $\tau : \mathcal{L} \mapsto \mathcal{S} \cup \{\epsilon\}$ maps a link to the signalized intersection from which vehicles enter the link. When $\tau(l) = \epsilon$ then link l is an entry link and no such upstream link exists; such a l is an entry link and $\mathcal{E} = \{l \in \mathcal{L} : \tau(l) = \epsilon\}$. Vehicles on l flow from the *upstream* intersection $\tau(l)$ to the *downstream* intersection $\sigma(l)$. An illustration of the notation is provided in Figure 7.1.

Each intersection uses a fixed time control strategy where a periodic sequence of non-conflicting links are allowed to discharge vehicles.

Assumption 7.2.1. *Each signalized intersection in the network has a common cycle period T or, identically, a common cycle frequency $\omega = \frac{2\pi}{T}$.*

Each link has an associated queue where vehicles wait to be released. We assume a fluid model for queues, which have a length $q_k(t) \in \mathbb{R}_{\geq 0}$ at time t and exhibit continuous dynamics driven by arrivals and departures for all $l \in \mathcal{L}$:

$$\dot{q}_l(t) = a_l(t) - d_l(t).$$

Link flows are controlled with red and green lights. Signals at all intersections obey a global clock and each intersection s has an *offset* $\theta_s \in [0, 2\pi]$ that represents a phase difference from that global clock. During each period, each link is actuated with a contiguous green subinterval. The middle of that subinterval is given by a *green split* $\gamma_l \in [0, 2\pi]$ and hence the middle of link l's green interval is actuated on $t = 2\pi n + \theta_{\sigma(l)} + \gamma_l$ for $n = 1, 2, \ldots$.

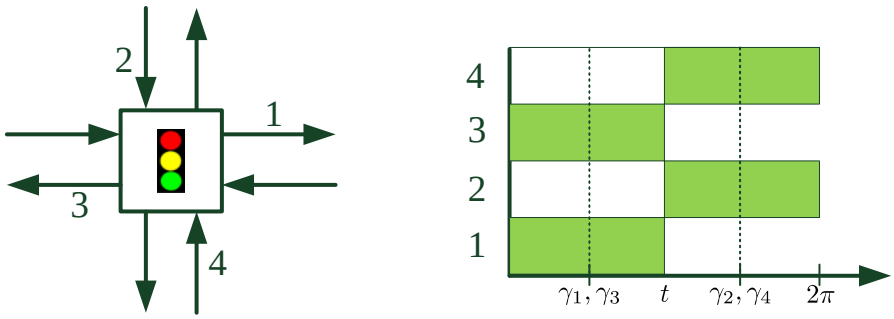


Figure 7.2: Green splits of a four way intersection.

Figure 7.2 shows a typical signal configuration for a four-way intersection, where the North-South movements and East-West movements are actuated on disjoint intervals. East-West links 1,3 have green splits $\gamma_1, \gamma_3 = \frac{\pi}{2}$ and North-South links 2,4 have green splits $\gamma_2, \gamma_4 = \frac{3\pi}{2}$.

Parameter $\beta_{kl} \in [0, 1]$ is the proportion of the vehicles exiting link k that enter link l. If A_k is the average flow of vehicles exiting link k, then the average flow of vehicles entering link l from k is $\beta_{kl} A_k$. Flows can only be non-zero on adjacent links; more precisely, $\beta_{kl} \neq 0$ only if $\tau(l) = \sigma(k)$. Furthermore, $\sum_{l \in \mathcal{L}} \beta_{kl} \leq 1$ where strict inequality signifies that some vehicles are exiting the network of modeled roads.

This framework is general enough to accommodate more granular models that include multiple lanes and unsignalized intersections such as stop signs or junctions. The microsimulation benchmark required a mild reformulation of the model above and further details are provided in Section 7.4.

Approximation with Sinusoids

For a network where all signalized intersections have a cycle time and the network is able to accomodate exogenous arrivals, the network is in a periodic steady state [73].

Assumption 7.2.2. *We assume that the network is under a periodic steady state with a period given by* $T = \frac{2\pi}{\omega}$*, i.e.,*

$$\forall l \in \mathcal{L} \qquad x_l(t) = x_l(t + Tn) \text{ for } n = 1, 2, \ldots. \tag{7.1}$$

The partitioning of a period into red and green subintervals induces a platooning effect where the departure rate of vehicles has peaks and troughs. This effect is approximated by modeling arrival and departure rates as sinusoids.

Entry Links

Entry links $l \in \mathcal{E}$ are assumed to have a periodic arrival rate $a_l(t)$

$$a_l(t) := A_l + \alpha_l \cos(\omega t - \varphi_l) \tag{7.2}$$

where A_l represents the mean arrival rate, α_l encodes a fluctuation, and φ_l is centered at the peak arrival rate. It's clear that $A_l, \alpha_l \geq 0$ and $A_l \geq \alpha_l$.

To obtain the departure process, first note that under the steady state assumption

$$\int_T a_l(t)dt = \int_T d_l(t)dt. \tag{7.3}$$

The departure rate $d_l(t)$ is then approximated with a sinusoid with a peak at the offset $\varphi_{\sigma(l)} + \gamma_l$.

$$d_l(t) := A_l(1 + \cos(\omega t - \varphi_{\sigma(l)} - \gamma_l)) \quad \forall l \in \mathcal{E}. \tag{7.4}$$

Internal Links

For non-entry link l the arrival rate is a sum of the departure rates of upstream links, except with an appropriate scaling factor from turning ratios and a travel time $\lambda_l \in [0, 2\pi]$ on link l

$$a_l(t) = \sum_{k \in \mathcal{L}} \beta_{kl} A_k (1 + \alpha_l \cos(\omega t - \theta_{\sigma(k)} - \gamma_k - \lambda_l)). \tag{7.5}$$

If $\tau(l) = \sigma(k)$ for any k, l such that $\beta_{kl} \neq 0$ then

$$a_l(t) = A_l + \alpha_l \cos(\omega t - \theta_{\tau(l)} - \varphi_l) \tag{7.6}$$

where α_l and φ_l for $l \in \mathcal{L} \setminus \mathcal{E}$ are

$$\alpha_l = \sqrt{\left(\sum_{k \in \mathcal{L}} \beta_{kl} A_k \cos(\gamma_k)\right)^2 + \left(\sum_{k \in \mathcal{L}} \beta_{kl} A_k \sin(\gamma_k)\right)^2} \tag{7.7}$$

$$\varphi_l = \lambda_l + \tan^{-1}\left(\frac{\sum_{k \in \mathcal{L}} \beta_{kl} A_k \sin(\gamma_k)}{\sum_{k \in \mathcal{L}} \beta_{kl} A_k \cos(\gamma_k)}\right). \tag{7.8}$$

Note that conservation of mass ensures that

$$\sum_{k \in \mathcal{L}} \beta_{kl} A_k = A_l. \tag{7.9}$$

The departure process is identical as for entry links:

$$d_l(t) := A_l(1 + \cos(\omega t - \varphi_{\sigma(l)} - \gamma_l)) \quad \forall l \in \mathcal{L}. \tag{7.10}$$

An internal queue's rate of change is given by

$$\dot{q}_l(t) = a_l(t) - d_l(t) \tag{7.11}$$

$$= \alpha_l \cos(\omega t - \theta_{\tau(l)} - \varphi_l) - A_l \cos(\omega t - \theta_{\sigma(l)} - \gamma_l). \tag{7.12}$$

Through a simple application of Euler's identity and phasors, the sum can be simplified to:

$$\dot{q}_l(t) = M_l \cos(\omega t - \eta_l)$$

for some phase shift η_l and amplitude:

$$M_l(\theta) = \sqrt{A_l^2 + \alpha_l^2 - 2A_l\alpha_l \cos\big((\theta_{\sigma(l)} + \gamma_l) - (\theta_{\tau(l)} + \varphi_l)\big)}. \tag{7.13}$$

The resulting explicit representation of $q_l(t)$ as a function of θ is:

$$q_l(t) = \frac{M_l(\theta)}{\omega} \cos(\omega t - \eta_l) + C_l \tag{7.14}$$

where $C_l > 0$ is a constant representing the average queue length with respect to time.

Optimization by Minimizing Queue Oscillation

For each link consider the cost function:

$$J_l(\theta) = \left(\frac{M_l(\theta)}{\omega}\right)^2$$

and observe from (7.13) that this shares a minimum with $R_l(\theta) = -A_l\alpha_l \cos\big((\theta_{\sigma(l)} + \gamma_l) - (\theta_{\tau(l)} + \varphi_l)\big)$. Both $J_l(\theta)$ and $R_l(\theta)$ are minimized when:

$$\theta_{\sigma(l)} + \gamma_l = \theta_{\tau(l)} + \varphi_l. \tag{7.15}$$

Condition (7.15) intuitively means that center of the green light $\theta_{\sigma(l)} + \gamma_l$ of intersection $\sigma(l)$ should match the time $\theta_{\tau(l)} + \varphi_l$ when the vehicle arrival rate from $\tau(l)$ peaks. Condition (7.15) is easily satisfied with a single link l and a pair of intersection offsets $\theta_{\sigma(l)}, \theta_{\tau(l)}$ but this is generally not the case for entire networks with complex topologies. Since such constraints cannot be satisfied simultaneously, we consider the sum $\sum_{l\in\mathcal{L}} R_l(\theta)$ as an objective function:

$$\min_{\theta} -\sum_{l\in\mathcal{L}} A_l\alpha_l \cos\big((\theta_{\sigma(l)} + \gamma_l) - (\theta_{\tau(l)} + \varphi_l)\big). \tag{7.16}$$

The A_l component of the coefficient weights prioritizes links with higher flow. The α_l component prioritizes links with higher arrival rates fluctuation; these links would benefit more from well aligned offsets.

7.3 Offset Optimization Algorithm

Reformulating as a Semidefinite Program

The offset optimization problem (7.16) is non-convex but is readily amenable to semi-definite relaxation. After introducing the substitution $z = (x_s, y_s)$ where $x_s = \cos(\theta_s)$ and $y_s = \sin(\theta_s)$, we can construct a W using trigonometric sum and difference identities such that the objective in (7.16) is equal to $z^T W z$. For $s, u \in \mathcal{S}$, let

$$W_1[s,u] = -\sum_{l \in \mathcal{L}_{s \to u}} A_l \alpha_l \cos(\varphi_l - \gamma_l)$$

and

$$W_2[s,u] = -\sum_{l \in \mathcal{L}_{s \to u}} A_l \alpha_l \sin(\varphi_l - \gamma_l)$$

where

$$\mathcal{L}_{s \to u} = \{l \in \mathcal{L} | \tau(l) = s \text{ and } \sigma(l) = u\}.$$

Finally, we arrive at W:

$$\bar{W} = \begin{bmatrix} W_1 & W_2 \\ -W_2 & W_1 \end{bmatrix}, W = \frac{1}{2}(\bar{W} + \bar{W}^T).$$

Each pair x_s, y_s must also satisfy $x_s^2 + y_s^2 = 1$. This constraint is encoded via a set of matrices M_s which essentially act as indicator functions that select elements of z. Let $E_s \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{S}|}$ have only one non-zero element such that $E_s[s,s] = 1$. Then,

$$M_s = \begin{bmatrix} E_s & 0 \\ 0 & E_s \end{bmatrix}.$$

The optimal value of problem (7.17) is equivalent to the original problem (7.16).

$$\begin{aligned} \min_{z \in \mathbb{R}^{2|\mathcal{S}|}} \; & z^T W z \\ s.t. \; & z^T M_s z = 1 \quad \forall s \in \mathcal{S}. \end{aligned} \tag{7.17}$$

Problem (7.17) is a non-convex quadratically constrained quadratic program (QCQP) that can be reformulated to a semidefinite program (SDP) with a rank constraint by using the circular property of the trace: $\mathbf{Tr}(z^T W z) = \mathbf{Tr}(W z z^T)$. Let $Z = z z^T$, we get

$$\begin{aligned} \min_{Z \succeq 0} \; & \mathbf{Tr}(WZ) \\ s.t. \; & \mathbf{Tr}(M_s Z) = 1 \quad \forall s \in \mathcal{S} \\ & \text{rank}(Z) = 1. \end{aligned} \tag{7.18}$$

The equivalent rank-constrained SDP formulation in (7.18) is a nonconvex problem due to its rank constraint. Dropping the rank constraint yields a convex optimization problem:

$$\begin{aligned} &\min_{Z \succeq 0} \mathbf{Tr}(WZ) \\ &s.t.\ \mathbf{Tr}(M_s Z) = 1 \quad \forall s \in \mathcal{S}. \end{aligned} \tag{7.19}$$

Because (7.19) is a relaxed convex SDP, the feasible set of problem (7.19) contains the feasible set of problem (7.18), and therefore, a solution to (7.19) provides a lower bound on (7.17) and when a minimizing Z is rank 1, then it is a tight lower bound.

Recovering a Feasible Solution from the SDP

For rank 1 solutions to the SDP we can transform back to the offsets via

$$\theta_s = \arctan\left(\frac{y_s}{x_s}\right) \quad \forall s \in \mathcal{S}. \tag{7.20}$$

In general, the solution to the semidefinite relaxation is not guaranteed to yield a rank one solution and be feasible for the original problem. One may consider taking the principal eigenvector of the minimizer $X = RR^T$ and feed it to equation (7.20); however, this is often not a local minimum, let alone a global minimum. A reasonable heuristic is to use the principal eigenvector as an initial point for our non-linear solver routine applied to the original offset optimization problem (7.16).

Burer-Monteiro Algorithm

Interior point solvers for semidefinite programs begin to run out of memory roughly when $|\mathcal{S}|$ is in the low thousands. The Burer-Monteiro algorithm for solving SDPs [11] applies the nonlinear change of variables $Z = RR^T$ where $R \in \mathbb{R}^{2|\mathcal{S}| \times r}$ has a rank upper bound of $r \leq 2|\mathcal{S}|$.

$$\begin{aligned} &\min_{R \in \mathbb{R}^{2|\mathcal{S}| \times r}} \mathbf{Tr}(WRR^T) \\ &s.t.\ \mathbf{Tr}(M_s RR^T) = 1 \quad \forall s \in \mathcal{S}. \end{aligned} \tag{7.21}$$

Although we pay the cost of transforming an originally convex optimization problem to a nonlinear program, we are free of conic constraints and dramatically reduce the dimension of the search space. Moreover, we are able to use memory-efficient first order and quasi-Newton methods.

Two key insights highlight the intuition behind the algorithm. First, for problems with compact search spaces there exists a global optimum of (7.19) of rank at most r^* where $\frac{r^*(r^*+1)}{2} \leq |\mathcal{S}|$ [5][84]. Thus, for $r^* \leq r$ the minima of (7.19) and (7.21) are equivalent. Second, it is commonly the case that a local minimum of (7.21) with $R \in \mathbb{R}^{2|\mathcal{S}| \times r}$ is either a global

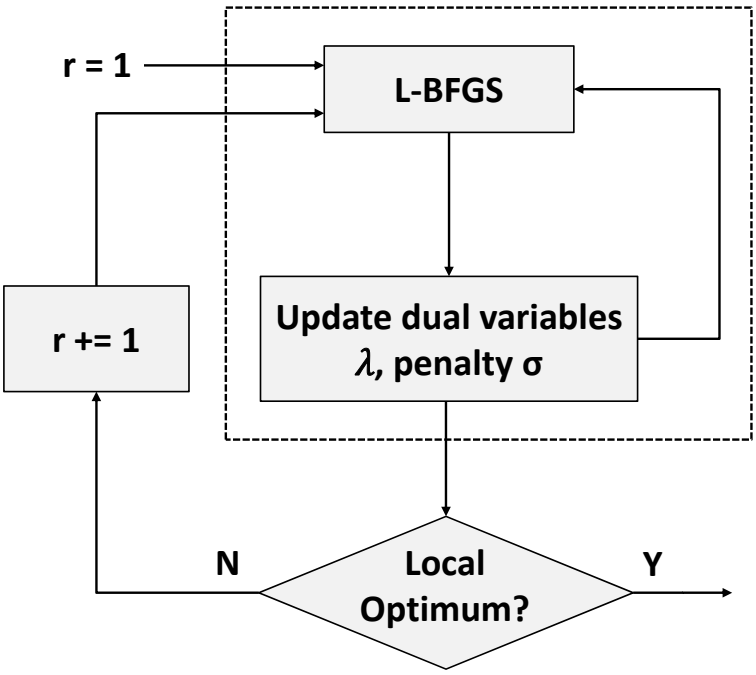


Figure 7.3: Control flowchart of the Burer-Monteiro Algorithm. The dashed box denotes the rank r subroutine in Algorithm 1.

minimum or can be escaped by increasing the rank of the search space to $R \in \mathbb{R}^{2|\mathcal{S}|\times(r+1)}$. Conditions for when this escape property holds have been a subject of theoretical interest; recent results [10] show that for SDPs where the solution set is a smooth manifold, the set of problematic matrices W where local minima do not have an escape is of measure 0. Both of these insights suggest that one can use the flowchart in Figure 7.3 to solve the SDP, where the dashed box encloses a subroutine to find local minima which we explain next.

Problem (7.21) is a equality constrained non-linear programming problem which we opt to solve with the method of multipliers [7]. The method of multipliers enforces the trace condition $\mathbf{Tr}(M_s RR^T) = 1$ with an augmented Lagrangian, which utilizes dual variables λ and penalty factor σ to penalize infeasibility:

$$L(R, \lambda, \sigma) = \mathbf{Tr}(WRR^T) - \sum_{s=1}^{|\mathcal{S}|} \lambda_s(\mathbf{Tr}(M_s RR^T) - 1) + \frac{\sigma}{2} \sum_{s=1}^{|\mathcal{S}|} (\mathbf{Tr}(M_s RR^T) - 1)^2. \quad (7.22)$$

For an appropriate (λ^*, σ^*) pair, a minimizer of (7.22) will recover a local minimizer of (7.21) [29]. To identify that pair (λ^*, σ^*), we use the iterative method shown in Algorithm 1. It first uses a nonlinear solver to minimize the augmented Lagrangian for a fixed (λ, σ), then it updates the dual variable λ while sufficient progress is being made and increases the penalty factor σ as soon as progress stalls. In our implementation, We opt to use the limited memory BFGS (L-BFGS) algorithm to minimize the augmented Lagrangian, but

$\sigma = \sigma_0, \lambda_s = \mathbf{0}$;
$k = 0$;
$v_0 = \frac{\sigma}{2} \sum_{s=1}^{|\mathcal{S}|} (\mathbf{Tr}(M_s(R^k)(R^k)^T) - 1)^2$;
while $|\nabla L(R, \lambda^k, \sigma^k)| > \epsilon$ *and* $\sigma_k < 2 \times 10^6$ **do**
  if $v \leq \eta v_k$ **then**
    $\lambda_i^{k+1} = \lambda_i^k - \sigma_k(\mathbf{Tr}(M_s(R^k)(R^k)^T) - 1)$ for all $s \in \mathcal{S}$;
    $\sigma_{k+1} = \sigma_k$;
    $v_{k+1} = v$;
  else
    $\lambda_i^{k+1} = y_i^k$ for all $s \in \mathcal{S}$;
    $\sigma_{k+1} = \gamma \sigma_k$;
    $v_{k+1} = v_k$;
  end
  $R^{k+1} =$ Nonlinear Solver$(R_{init} = R^k)$;
  $k = k + 1$
end

Algorithm 1: Rank 1 Burer-Monteiro Algorithm (method of multipliers) update for dual variable λ and penalty factor σ.

in principle any nonlinear solver that converges to a local minimum can be used. Parameter $\eta < 1$ specifies the rate of expected progress while $\gamma > 1$ specifies how quickly the penalty rate grows. In practice, we chose $\gamma = 2$ and $\eta = 0.9$ because much of the information about infeasibility was encoded in the dual variables λ. These parameters prevented ill-conditioning from the penalization factor growing too rapidly and ensured that the dual variable could successfully guide the local search to the feasible region. Once Algorithm 1 has terminated, a local minimum for (7.21) is obtained.

7.4 Numerical Experiments

Two examples of real-world networks that showcase the results of the Burer-Monteiro algorithm is presented in this section. The first example highlights the computational scalability of this problem while the second validates the model assumptions with a microscopic simulation. All benchmarks were performed on a 2013 Macbook Pro with 2.4GHz processor, 8GB memory, and with `python 2.7`.

The algorithms tested included "cold start" (random initialization) gradient descent (GD) for problem (7.16) before the coordinate transformation $z = (x, y)$. Lower bounds via the semidefinite relaxation are computed with `CVXOPT` [3] and the Burer-Monteiro method. The principal eigenvector from both of these algorithms are then used as "warm starts" (initialization with the solution of a preceding problem) for gradient descent. We also consider applying the method of multipliers (Rank 1 Burer-Monteiro algorithm) to problem (7.21)

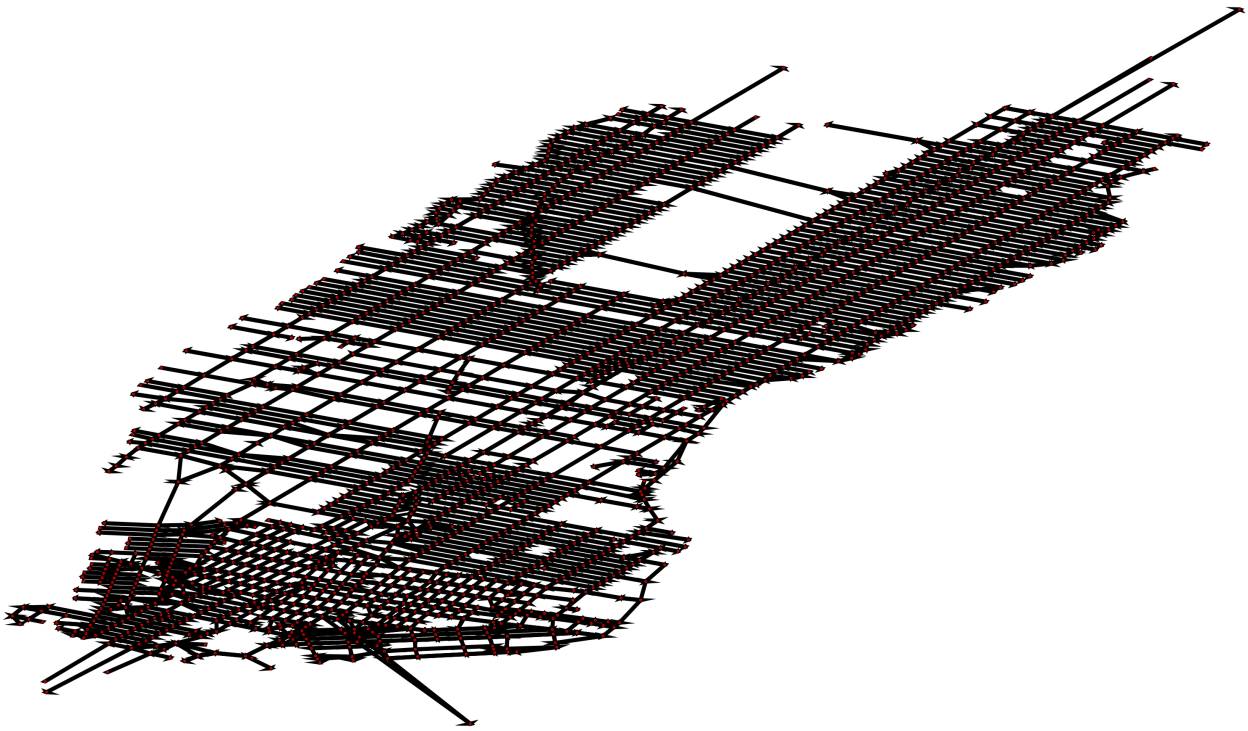


Figure 7.4: Depiction of optimized Manhattan graph.

without the rank update in the flowchart of Figure 7.3.

Manhattan Network

We demonstrate the scalability of the BM-algorithm by constructing a large synthetic network from `OpenStreetMap` data [76]. Our network is based off Manhattan in New York City, New York, USA and contains 1771 nodes and 3923 links, which is show in Figure 7.4 .

Green splits φ were assigned based on the orientation of the link as it enters the intersection in such a way that the green splits for North-South links are completely out of phase of East-West links. Travel times γ are proportional to link length. Turn ratios β between links were chosen so that a vehicle is more likely to travel straight than turn. Let $A_{\mathcal{E}}$ be a vector with $A_l = 0$ for internal links while A_l is given by (7.2) for $l \in \mathcal{E}$. The mean traffic flow is computed with $A = (I - \beta)^{-1} A_{\mathcal{E}} = \sum_{i=0}^{\infty} \beta^i A_{\mathcal{E}}$. A is finite because turn ratios by construction guarantee that β is a stochastic matrix.

Table 7.1 showcases the attained objective values and runtimes for a suite of algorithms. The Burer-Monteiro algorithm certifies a lower bound on the optimization problem while `CVXOPT` ran out of memory. Note that the warm start BM runtime is the sum of the BM runtime and the nonlinear solver with a warmstart. The Burer-Monteiro algorithm returned a result that is infeasible for the original problem because it has rank higher than 1. We use the eigenvector associated with the principal eigenvalue as a warm start for a gradient

Solver	Problem	Objective	Runtime (s)	Rank
Burer-Monteiro	(7.21)	-14895.66	347.63	3
CVXOPT	(7.19)	N/A	MemOut	N/A
Rank 1 Burer-Monteiro	(7.21)	-14888.02	145.02	1
Warm Start GD (BM)	(7.16)	-14887.26	421.25	1
Warm Start GD (CVXOPT)	(7.16)	N/A	MemOut	N/A
Cold Start GD	(7.16)	-14834.58	275.90	1

Table 7.1: Manhattan offset optimization benchmarks.

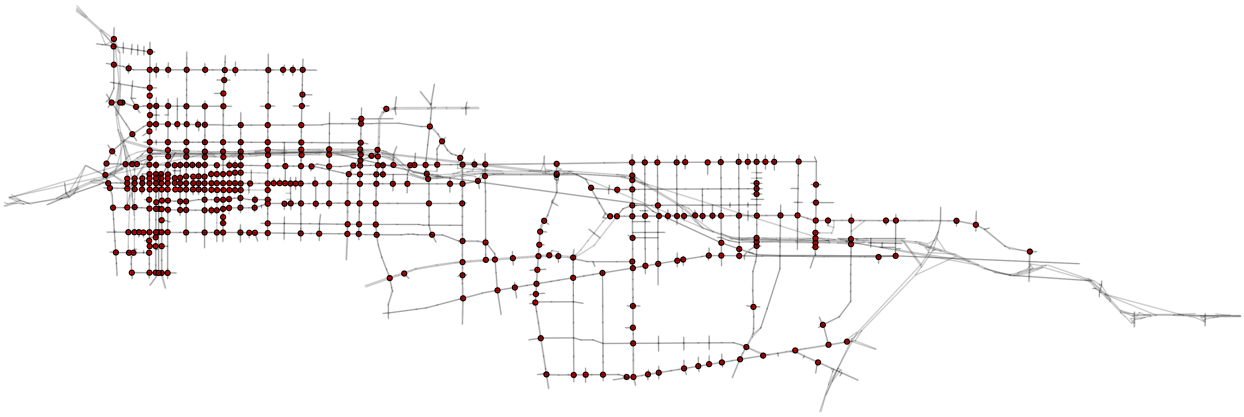


Figure 7.5: Aimsun network of Pasadena, CA, USA. The 420 optimized intersections are highlighted with a dark red circle.

descent algorithm. The cold start gradient descent method on the original problem performs remarkably well but the runtime is longer than even the Rank 1 Burer-Monteiro algorithm (method of multiplies). We attribute this to the amount of time it takes to compute a gradient for (7.16) which involves a large summation along the links $\mathcal{L}$ for each component.

Microsimulation of Pasadena, California

The results from our offset optimization algorithm are also validated with a microscopic simulation of a network in Pasadena, California, USA depicted in Figure 7.5. Simulations were run with Aimsun, which is a transportation modeling software used by traffic practitioners to model networks and evaluate control strategies. The network contains 1537 nodes, of which 459 are signalized intersections. This network is under ongoing development at the Connected Corridors Program at the U.C. Berkeley Institute of Transportation Studies [20]. The travel demands are obtained from area traffic studies between 2006 and 2014 and calibrated using in-built Aimsun routines. The geometric details of the network (lane counts, turn bays, etc) were manually verified against aerial imagery from Google Maps.

The microsimulation takes into consideration detailed parameters such as driver reaction

Cycle Time (s)	60	70	80	90	120
Nodes	85	59	138	59	79
Queues	759	552	1365	562	887

Table 7.2: Optimization subproblem size for each cycle time.

Solver	Problem	Objective	Total Runtime (s)	Rank
Pre-existing offset	N/A	-46561.95	N/A	N/A
Random offset	N/A	-2063.15	N/A	N/A
CVXOPT	(7.19)	-136767.71	3.14	188
Burer-Monteiro	(7.21)	-136767.70	0.004	5
Cold Start GD	(7.16)	-136265.80	51.55	1
Rank 1 Burer-Monteiro	(7.21)	-136767.70	0.004	1
Warm Start CVXOPT	(7.19)	-136767.71	11.07	1

Table 7.3: Pasadena, CA optimization benchmarks. Pre-existing and random offsets are included for comparison.

time, vehicle-to-vehicle gaps, and vehicle dynamics. The Aimsun model includes features such as unsignalized intersections and multiple lanes within a road [2]. Lanes can easily be accommodated by increasing the number of links in the network. The model in Section 7.2 assumes all intersections are signalized, so in the optimization problem all unsignalized intersections are ignored and the travel times between pairs of lane queues are updated to reflect the travel time of the shortest path between the queues plus any delay from stop signs.

There is not a single cycle time shared amongst all intersections, but 420 intersections have one of five cycle times: 60, 70, 80, 90, and 120 seconds. For each cycle time, the cost associated with a queue entering an intersection of a different cycle time are ignored. Table 7.2 shows how the problem size changes as different cycle times are considered. All right turn queues are ignored because vehicles are allowed to turn right with a red light. An optimization routine is then run on the smaller network with the modified cost function and optimization variables only over the intersections with the chosen cycle time. The minimizing offsets for each cycle time are then concatenated to create a full 420 dimension offset vector.

All offset optimization subproblems took a few seconds and the summation of the objectives, runtimes, and rank for each of the five sub-problems is shown in Table 7.3. The Burer-Monteiro method recovers the same value as `CVXOPT`, but with a feasible, rank 1 objective each time. Like the Manhattan network, the output from gradient descent with a random initialization is nearly optimal. We also included random offsets and pre-existing offsets from the Aimsun model as a comparison.

To validate the assumptions involved in the optimization formulation, we collected simulations metrics for the same random, pre-existing, and optimized (via the method of mul-

Metric	Random	Baseline	Rank 1 Burer-Monteiro	Improvement
Stop time $\left(\frac{\text{s}}{\text{veh}\cdot\text{hr}}\right)$	12.33	12.01	11.57	3.65%
Queue (veh)	348.8	337.4	327.9	2.80%
Flow (veh/hr)	169.9	170.5	171.4	0.5%

Table 7.4: Metrics for different offsets. Each value is a mean among all queues that are controlled by an intersection with an optimized offset. The improvement column compares the optimized against the baseline offsets.

tipliers) offsets that appear in Table 7.3. Table 7.4 shows that both the delay time and the mean-queue length improve for the optimized versus the baseline version, despite the fact that there was a small net increase in the average flow experienced by the signalized intersections.

7.5 Summary

This study demonstrates the Burer-Monteiro method to the offset optimization problem for a large network of signalized intersections. The Manhattan example showcases the scalability of the BM algorithm. Meanwhile for the Pasadena, CA network a comparison among a suite of algorithms demonstrates that the BM method reliably finds a global optimum of the semidefinite relaxation. A detailed microsimulation of Pasadena, California demonstrates that the optimization routine yields tangible improvements in delay time and mean queue occupancy, in spite of modeling assumptions such as the global cycle time, fluid queue model, sinusoidal approximation, and absence of congestion. The speedup offered by the BM method raises the intriguing possibility of real-time coordination of large traffic networks.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

This dissertation covered topics related to traffic flow estimation, traffic modeling, and control techniques by field test implementation. Two freeway control methods, coordinated ramp metering and variable speed advisory, have been implemented and evaluated on their respective chosen test sites. We also presented offset optimization of intersection networks, that can be used for signal design of a large scale network.

In Chapter 2, we reviewed previous research related to the materials discussed in this dissertation, introducing traffic data collection technologies, traffic flow models, and commonly used traffic controls.

In Chapter 3, we proposed a boundary flow prediction method that combines the most recent traffic data with historical traffic data. An autoregressive moving average with exogenous input (ARMAX) was estimated on-line based on the most recent vehicle detector station (VDS) data. An optimal multiple step ahead traffic demand predictor is obtained based on the estimated ARMAX model by solving a corresponding Bezout equation for each predictor. Results obtained using empirical freeway mainline and on-ramp data show that this method outperforms methods that rely only on the historical average of the data to perform a prediction, especially during days with unusual traffic flow demands, such as a Super Bowl Sunday. Due to its simplicity and robustness, this method is useful in practical applications.

In Chapter 4, we presented a procedure for constructing and calibrating a microsimulation model of a congested freeway with multiple vehicle classes using AIMSUN. This procedure was applied to a 13 mile long stretch of SR-99 Northbound in Sacramento, California. The test site includes a mainline high occupancy vehicle (HOV) lane, 16 metered on-ramps with and without HOV lanes, freeway interconnection and three interacting bottlenecks. The model construction and calibration procedure is as follows: (1) building the road geometry in AIMSUN, (2) collection of the traffic condition data from the PeMS database, (3) imputation of missing data, (4) estimation of on-ramp demand and off-ramp turning ratio,

(5) identification of recurring bottlenecks, (6) setting model parameters, and (7) adjustment of model parameters such that the simulation results match with the field measurements. Model calibration results indicated that the simulated flow match with the field data near bottleneck locations reasonably well under the criterion of root mean square percentage error (RMSPE) and GEH criterion. The calibrated AIMSUN microsimulation model was subsequently used as a simulation test bed for implementing a coordinated ramp metering (CRM) control strategy in Chapter 5, which has the goal of alleviating the heavy congestion of the SR-99N freeway during the morning peak period.

In Chapter 5, we presented the field implementation and testing of a coordinated ramp metering (CRM) algorithm based on a simplified optimal control approach. The test site was the California State Route 99 Northbound (SR-99N) corridor in Sacramento between Calvine Road and the SR-50 interchange after 12th Avenue (Abs. PostMile 290.454 - 299.467). It is a 9-mile long corridor with 11 on-ramps, to which the CRM algorithm has been applied. We compared the vehicle-miles-travelled (VMT), vehicle-hours-travelled (VHT), and the ratio VMT/VHT (defined as system efficiency in PeMS and interpreted as average speed) during field tests in 2016 with data from 2015 in the same period, during the morning peak hours (6:00 AM-9:00 AM), VMT/VHT and verified an increase of 7.25% on average, indicating a significant traffic efficiency improvement. During the evening peak hours (3:00 PM-6:00 PM), VMT/VHT decreased by 0.44% on average, indicating no traffic efficiency improvement. The reason for the lack of improvement during the evening commute was that the traffic was not heavily congested during most evening hours. Test results suggest that the CRM algorithm tested could be more effective for heavily congested traffic. The system has been deployed for daily operation along the SR-99N corridor after the test period.

In Chapter 6, we presented the field implementation and testing of variable speed advisory (VSA) for bottleneck flow maximization based on speed and occupancy measurement. The test site was at the California State Route 78 Eastbound (SR-78E) corridor in San Diego, between Vista Village Drive in the City of Vista and U.S. Route 15 (US-15) interchange in the city of Escondido (Abs. PostMile 6.32 - 17.73). The test site is an 11-mile long corridor with 10 on-ramps and 10 off-ramps to which the VSA algorithm has been applied. By comparing the vehicle miles-travelled (VMT), vehicle-hours-travelled (VHT), and the ratio VMT/VHT (freeway efficiency) of PeMS data which is independent of the data from VSA system for objectivity in performance evaluation, the results for before and after VSA control implementation were analyzed in depth. During the morning peak hours (6:00 AM-9:00 AM), VMT/VHT increased by 8.71% on average, indicating significant traffic efficiency improvement. During the evening peak hours (2:00 PM-7:00 PM), VMT/VHT increased only by 2.8% on average, which suggests a slight traffic efficiency improvement. The test results suggested that the proposed VSA algorithm can be deployed for heavily congested traffic to alleviate congestion. Findings of this field test will help to further develop VSA control algorithms and improve VSA traffic operations benefits.

In Chapter 7, we studied an arterial signalization optimization problem, which seeks to coordinate traffic signals in a large traffic network. By assuming that all signals have a common cycle time and that the arrival/departure rates at each intersection can be approxi-

mated by sinusoids, the original non-convex offset optimization problem can be relaxed into a semidefinite program (SDP). SDP solvers unfortunately run out of memory for larger networks with thousands of intersections. We used the Burer-Monteiro (BM) method for solving large SDPs, which avoids conic constraints and solves a lower dimensional problem, at the expense of being non-convex. A synthetic New York City example with 1771 intersections showcases the scalability of the BM method. Another example involving 420 intersections in Los Angeles demonstrates that the BM method recovers optimal solutions of the SDP. Moreover, a detailed microsimulation of the Los Angeles network showed that the optimized offsets result in reduced delay time and smaller queues.

8.2 Future Work

There are multiple possible directions for future research related to the topics studied in this dissertation. First, for the traffic flow prediction developed in Chapter 3, we can further study the prediction of an error bound and the minimization of the prediction error, which provides information to the traffic flow forecast engine about the reliability of the boundary forecast. Also, to incorporate route choice of travelers before they enter the network, we can investigate spatial correlations of neighboring on-ramp detectors, which is expected to increase the overall accuracy and decrease the uncertainty of the prediction. Second, the model calibration procedure presented in Chapter 4 uses historical data to build up a simulation test platform for preparation of field implementation test. It is worth to develop a microscopic freeway simulator with real-time traffic data and automatic calibration if more detailed data about driver's behavior are available . Third, due to the limitation of the test site, the coordinated ramp metering (CRM) field implementation test in Chapter 5 and the variable speed advisory (VSA) field implementation test in Chapter 6 are studied independently in different test sites. It is possible to develop a combined traffic management strategy including the coordination of ramp metering, variable speed advisory and intersection signals. Fourth, for the study in Chapter 7, we can extend the intersection model by given traffic flow prediction such that it can accommodate time varying demands with stochastic uncertainty. Finally, future's intelligent transportation systems will be equipped with a heterogeneous of sensors both within the infrastructure and with connected or automated vehicles, which increases the complexity of modeling and control of existing traffic network. One can seek for improvement of traditional traffic management methods to accommodate future's complex traffic network by considering the presence of connected and automated vehicles.

Bibliography

[1] *AIMSUN 7 Dynamic Simulators Users Manual*. 2013 (page 37).

[2] Aimsun https://www.aimsun.com/. accessed 1-May-2019 (pages 12, 29, 102).

[3] M. S. Andersen, J. Dahl, and L. Vandenberghe. *Cvxopt: A python package for convex optimization, version 1.1. 6*. 2013 (page 99).

[4] K. Ashok. "Estimation and prediction of time-dependent origin-destination flows". PhD thesis. Massachusetts Institute of Technology, 1996 (page 15).

[5] A. I. Barvinok. "Problems of distance geometry and convex properties of quadratic maps". In: *Discrete & Computational Geometry* 13.2 (1995), pp. 189–202. ISSN: 1432-0444 (page 97).

[6] R. Bertini, S. Boice, and K. Bogenberger. "Dynamics of Variable Speed Limit System Surrounding Bottleneck on German Autobahn". In: *Transportation Research Record: Journal of the Transportation Research Board* 1978 (2006), pp. 149–159 (page 65).

[7] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 2014 (page 98).

[8] N. Bhouri, H. Haj-Salem, and J. Kauppila. "Isolated versus coordinated ramp metering: Field evaluation results of travel time reliability and traffic impact". In: *Transportation Research Part C: Emerging Technologies* 28 (2013). Euro Transportation: selected paper from the EWGT Meeting, Padova, September 2009, pp. 155 –167. ISSN: 0968-090X (page 43).

[9] M. Bierlaire. "Simulation and optimization: A short review". In: *Transportation Research Part C: Emerging Technologies* 55.0 (2015), 4 –13, Engineering and Applied Sciences Optimization (OPT–i) –Professor Matthew G. Karlaftis Memorial Issue. ISSN: 0968-090X (page 29).

[10] N. Boumal, V. Voroninski, and A. S. Bandeira. "The Non-convex Burer–Monteiro Approach Works on Smooth Semidefinite Programs". In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. NIPS'16. Barcelona, Spain: Curran Associates Inc., 2016, pp. 2765–2773. ISBN: 978-1-5108-3881-9 (page 98).

[11] S. Burer and R. D. C. Monteiro. "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization". In: *Mathematical Programming* 95 (2003), pp. 329–357 (pages 91, 97).

[12] E. F. Camacho and C. A. Bordons. *Model Predictive control*. Springer London, 2007 (page 50).

[13] M. J. Cassidy. "Freeway On-Ramp Metering, Delay Savings, and Diverge Bottleneck". In: *Transportation Research Record* 1856.1 (2003), pp. 1–5 (page 12).

[14] M. Cassidy. "Handbook of Transportation Science". In: ed. by R. Hall. Springer Science & Business Media, 2012. Chap. 6 Traffic Flow and Capacity (page 72).

[15] C. Chen, Z. Jia, and P. Varaiya. "Causes and Cures of Highway Congestion". In: *IEEE Control Systems* 21.6 (2001), pp. 26–32. ISSN: 1066-033X (page 58).

[16] C. Chen et al. "Freeway Performance Measurement System: Mining Loop Detector Data". In: *Transportation Research Record* 1748.1 (2001), pp. 96–102 (page 8).

[17] D. Chen, S. Ahn, and A. Hegyi. "Variable speed limit control for steady and oscillatory queues at fixed freeway bottlenecks". In: *Transportation Research Part B: Methodological* 70 (2014), pp. 340 –358. ISSN: 0191-2615 (page 13).

[18] A. Chow et al. "An empirical study of traffic breakdown for Variable Speed Limit and Ramp Metering control". In: *89th TRB Annual Meeting*. 2009, pp. 10–14 (page 49).

[19] B. Ciuffo and V. Punzo. ""No Free Lunch"; Theorems Applied to the Calibration of Traffic Simulation Models". In: *Intelligent Transportation Systems, IEEE Transactions on* 15.2 (2014), pp. 553–562. ISSN: 1524-9050 (page 29).

[20] Connected Corridors Program https://connected-corridors.berkeley.edu/. accessed 1-May-2019 (pages 14, 91, 101).

[21] S. Coogan et al. "Offset optimization for a network of signalized intersections via semidefinite relaxation". In: 2015 (pages 91, 92).

[22] A. Dabiri and B. Kulcsár. "Distributed Ramp Metering–A Constrained Discharge Flow Maximization Approach". In: *IEEE Transactions on Intelligent Transportation Systems* PP.99 (2017), pp. 1–14. ISSN: 1524-9050 (page 44).

[23] C. F. Daganzo. "The Cell Transmission Model: A dynamic representation of highway traffic consistent with the hydrodynamic theory". In: *Transportation Research Part B: Methodological* 28.4 (1994), pp. 269 –287. ISSN: 0191-2615 (pages 9, 10, 14).

[24] G. Dervisoglu et al. "Automatic Calibration of the Fundamental Diagram and Empirical Observations on Capacity". In: *Transportation Research Board, 88th Annual Meeting of the Transportation Research Board, Washington, D.C.* (Feb. 2009) (page 10).

[25] T. Djukic, J. W. C. Van Lint, and S. P. Hoogendoorn. "Exploring Application Perspectives of Principal Component Analysis in Predicting Dynamic Origin-Destination Matrices". In: *Transportation Research Board 91st Annual Meeting*. 12-2702. 2012 (page 15).

[26] R. Dowling et al. "Guidelines for Calibration of Microsimulation Models: Framework and Applications". In: *Transportation Research Record: Journal of the Transportation Research Board* 1876 (2004), pp. 1–9 (page 28).

[27] L. Faulkner et al. "Evaluation of HERO-Coordinated Ramp Metering Installation at M1 and M3 Freeways in Queensland, Australia". In: *Transportation Research Record: Journal of the Transportation Research Board* 2470 (2014), pp. 13–23 (pages 13, 43).

[28] FHWA. "Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software". In: *Publication No. FHWA-HRT-04-040* (2004) (pages 28, 37).

[29] R. Fletcher. *Practical Methods of Optimization; (2Nd Ed.)* New York, NY, USA: Wiley-Interscience, 1987. ISBN: 0-471-91547-5 (page 98).

[30] G. Gomes and R. Horowitz. "Optimal freeway ramp metering using the asymmetric cell transmission model". In: *Transportation Research Part C: Emerging Technologies* 14.4 (2006), pp. 244 –262. ISSN: 0968-090X (page 9).

[31] G. Gomes, A. May, and R. Horowitz. "Congested Freeway Microsimulation Model Using VISSIM". In: *Transportation Research Record: Journal of the Transportation Research Board* 1876 (2004), pp. 71–81 (page 29).

[32] G. C. Goodwin and K. S. Sin. *Adaptive Filtering Prediction and Control*. Dover Publications, 2009 (pages 19–21).

[33] Google Maps https://www.google.com/maps. accessed 1-May-2019 (page 30).

[34] B. Greenshields et al. "A study of traffic capacity". In: *Proceedings of the 14th annual meeting of the Highway Research Board* (1935), pp. 448–477. ISSN: 0096-1027 (page 9).

[35] S. Gunnarsson. "Combining tracking and regularization in recursive least squares identification". In: *Decision and Control, 1996., Proceedings of the 35th IEEE Conference on*. Vol. 3. 1996, 2551–2552 vol.3 (pages 20, 22).

[36] D. K. Hale et al. "Optimization-based Assisted Calibration of Traffic Simulation Models". In: *Transportation Research Part C: Emerging Technologies* 55.0 (2015), 100 – 115, Engineering and Applied Sciences Optimization (OPT–i) –Professor Matthew G. Karlaftis Memorial Issue. ISSN: 0968-090X (page 29).

[37] B. Harbord et al. "A Flexible Approach to Motorway Control". In: *13th World Congress on Intelligent Transport Systems*. 2006 (page 65).

[38] A. Hegyi and S. P. Hoogendoorn. "Dynamic speed limit control to resolve shock waves on freeways - Field test results of the SPECIALIST algorithm". In: *13th International IEEE Conference on Intelligent Transportation Systems*. 2010, pp. 519–524 (pages 13, 65).

[39] Y. Hollander and R. Liu. "The Principles of Calibrating Traffic Microsimulation Models". English. In: *Transportation* 35.3 (2008), pp. 347–362. ISSN: 0049-4488 (page 28).

[40] S. Hoogendoorn et al. "Assessment of Dynamic Speed Limits on Freeway A20 near Rotterdam, Netherlands". In: *Transportation Research Record: Journal of the Transportation Research Board* 2380 (2013), pp. 61–71 (page 65).

[41] S. P. Hoogendoorn and P. H. L. Bovy. "State-of-the-art of vehicular traffic flow modelling". In: *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering* 215.4 (2001), pp. 283–303 (page 9).

[42] Z. Jia et al. "The PeMS algorithms for accurate, real-time estimates of g-factors and speeds from single-loop detectors". In: *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585)*. 2001, pp. 536–541 (page 8).

[43] X. Jiang and H. Adeli. "Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting". In: *Journal of Transportation Engineering* 131.10 (2005), pp. 771–779 (page 16).

[44] Y. Kamarianakis and P. Prastacos. "Space-time modeling of traffic flow". In: *Computers & Geosciences* 31.2 (2005), pp. 119 –133. ISSN: 0098-3004 (page 16).

[45] Y. Kamarianakis, W. Shen, and L. Wynter. "Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO". In: *Applied Stochastic Models in Business and Industry* 28.4 (2012), pp. 297–315. ISSN: 1526-4025 (page 16).

[46] Y. Kan et al. "Local ramp metering with distant downstream bottlenecks: A comparative study". In: *Transportation Research Part C: Emerging Technologies* 62 (2016), pp. 149 –170. ISSN: 0968-090X (page 43).

[47] M. Karlaftis and E. Vlahogianni. "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights". In: *Transportation Research Part C: Emerging Technologies* 19.3 (2011), pp. 387 –399. ISSN: 0968-090X (page 16).

[48] J. Kianfar, P. Edara, and C. Sun. "Operational Analysis of a Freeway Variable Speed Limit System in St. Louis, Missouri". In: *Journal of Intelligent Transportation Systems* 19.4 (2015), pp. 385–398 (page 65).

[49] L. A. Klein, M. K. Mills, and D. R. Gibson. *Traffic Detector Handbook: Third Edition*. Tech. rep. Federal Highway Administration, 2006 (page 7).

[50] V. Knoop, S. P. Hoogendoorn, and H. van Zuylen. "Empirical Differences Between Time Mean Speed and Space Mean Speed". In: *Traffic and Granular Flow '07*. Ed. by C. Appert-Rolland et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 351–356. ISBN: 978-3-540-77074-9 (page 48).

[51] A. Kotsialos et al. "Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool METANET". In: *IEEE Transactions on Intelligent Transportation Systems* 3.4 (2002), pp. 282–292. ISSN: 1524-9050 (page 9).

[52] D. Kress, R. L. Bertini, and K. Bogenberger. "Implementations of Variable Speed Limit Systems: A Worldwide Overview". In: *Transportation Research Board 97th Annual Meeting*. 2018 (page 65).

[53] E. Kwon et al. "Minnesota Variable Speed Limit System: Adaptive Mitigation of Shock Waves for Safety and Efficiency of Traffic Flows". In: *Transportation Research Board 90th Annual Meeting*. 2011 (page 65).

[54] S. Lee and D. B. Fambro. "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting". In: *Transportation Research Record: Journal of the Transportation Research Board* 1678.1 (1999), pp. 179–188 (page 15).

[55] D. Li and P. Ranjitkar. "A fuzzy logic-based variable speed limit controller". In: *Journal of Advanced Transportation* 49.8 (2015), pp. 913–927 (page 65).

[56] D. Li, P. Ranjitkar, and A. Ceder. "A Logic Tree Based Algorithm for Variable Speed Limit Controllers To Manage Recurrently Congested Bottlenecks". In: *Transportation Research Board 93rd Annual Meeting*. 2014 (page 65).

[57] P. Y. Li et al. "An automated highway system link layer controller for traffic flow stabilization". In: *Transportation Research Part C: Emerging Technologies* 5.1 (1997), pp. 11–37 (page 69).

[58] Z. Li et al. "Development of a Control Strategy of Variable Speed Limits to Reduce Rear-End Collision Risks Near Freeway Recurrent Bottlenecks". In: *IEEE Transactions on Intelligent Transportation Systems* 15.2 (2014), pp. 866–877. ISSN: 1524-9050 (page 13).

[59] M. J. Lighthill and G. B. Whitham. "On kinematic waves II. A theory of traffic flow on long crowded roads". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 229.1178 (1955), pp. 317–345 (page 9).

[60] M. Lippi, M. Bertini, and P. Frasconi. "Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning". In: *Intelligent Transportation Systems, IEEE Transactions on* 14.2 (2013), pp. 871–882. ISSN: 1524-9050 (page 16).

[61] L. Ljung. *System identification - Theory for the User*. Prentice-Hall, 1999 (page 20).

[62] X.-Y. Lu and S. E. Shladover. "Review of Variable Speed Limits and Advisories: Theory, Algorithms, and Practice". In: *Transportation Research Record* 2423.1 (2014), pp. 15–23 (page 65).

[63] X.-Y. Lu and A. Skabardonis. "Freeway Traffic Shockwave Analysis: Exploring the NGSIM Trajectory Data". In: *Transportation Research Board 86th Annual Meeting*. 2007 (page 72).

[64] X.-Y. Lu et al. "Estimating Traffic Speed with Single Inductive Loop Event Data". In: *Transportation Research Record: Journal of the Transportation Research Board* 2308 (2012), pp. 157–166 (page 46).

[65] X.-Y. Lu et al. "Freeway Micro-simulation Calibration: Case Study Using AIMSUN and VISSIM with Detailed Field Data". In: *Transportation Research Board, 93rd Annual Meeting of the Transportation Research Board, Washington, D.C.,* (2014) (page 29).

[66] X.-Y. Lu et al. "Novel Freeway Traffic Control with Variable Speed Limit and Coordinated Ramp Metering". In: *Transportation Research Record: Journal of the Transportation Research Board* 2229 (2011), pp. 55–65 (pages 28, 44, 46).

[67] J. Ma, H. Dong, and H. Zhang. "Calibration of Microsimulation with Heuristic Optimization Methods". In: *Transportation Research Record: Journal of the Transportation Research Board* 1999 (2007), pp. 208–217 (page 29).

[68] L. E. Y. Mimbela and L. A. Klein. *Summary of Vehicle Detection and Surveillance Technologies Used in Intelligent Transportation Systems*. Tech. rep. Federal Highway Administration, U.S. Department of Transportation, 2000 (page 7).

[69] Minnesota Department of Transportation Smart Lanes http://www.dot.state.mn.us/smartlanes/. accessed 1-Jul-2018 (page 65).

[70] E. R. Mller et al. "Microsimulation Analysis of Practical Aspects of Traffic Control With Variable Speed Limits". In: *IEEE Transactions on Intelligent Transportation Systems* 16.1 (2015), pp. 512–523. ISSN: 1524-9050 (page 13).

[71] A. Muralidharan. "Tools for modeling and control of freeway networks". PhD thesis. University of California, Berkeley, 2012 (page 15).

[72] A. Muralidharan and R. Horowitz. "Imputation of Ramp Flow Data for Freeway Traffic Simulation". In: *Transportation Research Record: Journal of the Transportation Research Board* 2099 (2009), pp. 58–64 (pages 15, 45).

[73] A. Muralidharan, R. Pedarsani, and P. Varaiya. "Analysis of fixed-time control". In: *Transportation Research Part B: Methodological* 73 (2015), pp. 81–90. ISSN: 0191-2615 (page 93).

[74] Next Generation Simulation (NGSIM) https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm. accessed 1-May-2019 (page 8).

[75] A. Nissan and H. N. Koutsopoulosb. "Evaluation of the Impact of Advisory Variable Speed Limits on Motorway Capacity and Level of Service". In: *Procedia - Social and Behavioral Sciences* 16 (2011). 6th International Symposium on Highway Capacity and Quality of Service, pp. 100 –109. ISSN: 1877-0428 (page 13).

[76] OpenStreetMap https://wiki.openstreetmap.org/. accessed 1-May-2019 (pages 30, 100).

[77] G Paesani et al. "System wide adaptive ramp metering in southern California". In: *ITS America 7th Annual Meeting*. 1997 (page 13).

[78] M. Papageorgiou et al. "Review of road traffic control strategies". In: *Proceedings of the IEEE* 91.12 (2003), pp. 2043–2067. ISSN: 0018-9219 (page 12).

[79] M. Papageorgiou, J.-M. Blosseville, and H. Haj-Salem. "Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part II: Coordinated on-ramp metering". In: *Transportation Research Part A: General* 24.5 (1990), pp. 361 –370. ISSN: 0191-2607 (page 13).

[80] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville. "ALINEA: a local feedback control law for on-ramp metering". In: *Transportation Research Record* 1320 (1991), pp. 58–64 (page 13).

[81] I. Papamichail et al. "Coordinated ramp metering for freeway networks A model-predictive hierarchical control approach". In: *Transportation Research Part C: Emerging Technologies* 18.3 (2010). 11th IFAC Symposium: The Role of Control, pp. 311 –331. ISSN: 0968-090X (page 44).

[82] PARAMICS https://www.paramics.co.uk/en/. accessed 1-May-2019 (page 12).

[83] C. Pasquale et al. "Two-class freeway traffic regulation to reduce congestion and emissions via nonlinear optimal control". In: *Transportation Research Part C: Emerging Technologies* 55 (2015). Engineering and Applied Sciences Optimization (OPT-i) - Professor Matthew G. Karlaftis Memorial Issue, pp. 85 –99. ISSN: 0968-090X (page 58).

[84] G. Pataki. "On the Rank of Extreme Matrices in Semidefinite Programs and the Multiplicity of Optimal Eigenvalues". In: *Mathematics of Operations Research* 23.2 (1998), pp. 339–358. ISSN: 0364765X, 15265471 (page 97).

[85] PATH http://www.path.berkeley.edu. accessed 1-May-2019 (pages 44, 75).

[86] PeMS http://pems.dot.ca.gov/. accessed 1-May-2019 (pages 8, 15, 16, 21, 30, 45, 46, 49, 55, 58, 85).

[87] W. H. Press et al. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. 3rd ed. New York, NY, USA: Cambridge University Press, 2007. ISBN: 0521880688, 9780521880688 (page 50).

[88] V. Punzo, M. Montanino, and B. Ciuffo. "Do We Really Need to Calibrate All the Parameters? Variance-Based Sensitivity Analysis to Simplify Microscopic Traffic Flow Models". In: *Intelligent Transportation Systems, IEEE Transactions on* 16.1 (2015), pp. 184–193. ISSN: 1524-9050 (page 29).

[89] B. Recht, M. Fazel, and P. Parrilo. "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization". In: *SIAM Review* 52.3 (2010), pp. 471–501 (page 91).

[90] H. Remeijn. "The Dutch Motorway Control System 13 Years of Evolution". In: *IFAC Proceedings Volumes* 27.12 (1994). IFAC Symposium on Transportation Systems: Theory and Application of Advanced Technology, Tianjin, PRC, 24-26 August, pp. 749 –754. ISSN: 1474-6670 (page 65).

[91] P. I. Richards. "Shock Waves on the Highway". In: *Operations research* 4.1 (1956), pp. 42–51 (page 9).

[92] G. Riggins et al. "Evaluation of driver compliance to displayed variable advisory speed limit systems: comparison between Germany and the US". In: *Transportation Research Procedia* 15 (2016), pp. 640–651 (page 65).

[93] T. V. da Rocha et al. "Does Traffic-related Calibration of Car-following Models Provide Accurate Estimations of Vehicle Emissions?" In: *Transportation Research Part D: Transport and Environment* 34 (2015), pp. 267 –280. ISSN: 1361-9209 (page 29).

[94] D. Schrank et al. *2015 Urban Mobility Scorecard*. Tech. rep. Texas A&M Transportation Institute, 2015 (page 1).

[95] T. Schreiter, J. W. C. Van Lint, and S. P. Hoogendoorn. "Vehicle-class Specific Control of Freeway Traffic". In: *Transportation Research Board 92nd Annual Meeting*. 13-0585. 2013 (page 14).

[96] S. Shekhar and B. M. Williams. "Adaptive seasonal time series models for forecasting short-term traffic flow". In: *Transportation Research Record: Journal of the Transportation Research Board* 2024.1 (2008), pp. 116–125 (page 16).

[97] A. Singer. "Angular synchronization by eigenvectors and semidefinite programming". In: *Applied and Computational Harmonic Analysis* 30.1 (2011), pp. 20 –36. ISSN: 1063-5203 (page 91).

[98] E. Smaragdis and M. Papageorgiou. "Series of New Local Ramp Metering Strategies". In: *Transportation Research Record* 1856.1 (2003), pp. 74–86 (page 13).

[99] B. L. Smith, B. M. Williams, and R. K. Oswald. "Comparison of parametric and nonparametric models for traffic flow forecasting". In: *Transportation Research Part C: Emerging Technologies* 10.4 (2002), pp. 303 –321. ISSN: 0968-090X (page 16).

[100] S. Smulders. "Control of freeway traffic flow by variable speed signs". In: *Transportation Research Part B: Methodological* 24.2 (1990), pp. 111 –132. ISSN: 0191-2615 (page 65).

[101] F. Soriguera, J. M. Torn, and D. Rosas. "Assessment of Dynamic Speed Limit Management on Metropolitan Freeways". In: *Journal of Intelligent Transportation Systems* 17.1 (2013), pp. 78–90 (page 13).

[102] F. Soriguera et al. "Effects of low speed limits on freeway traffic flow". In: *Transportation Research Part C: Emerging Technologies* 77 (2017), pp. 257 –274. ISSN: 0968-090X (page 65).

[103] Traffic Logix https://trafficlogix.com. accessed 1-May-2019 (page 75).

[104] TSIS-CORSIM https://mctrans.ce.ufl.edu/mct/index.php/tsis-corsim/. accessed 1-May-2019 (page 12).

[105] V. Vaze et al. "Calibration of Dynamic Traffic Assignment Models with Point-to-Point Traffic Surveillance". In: *Transportation Research Record: Journal of the Transportation Research Board* 2090 (2009), pp. 1–9 (page 36).

[106] VISSIM http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/. accessed 1-May-2019 (page 12).

[107] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias. "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach". In: *Transportation Research Part C: Emerging Technologies* 13.3 (2005), pp. 211 –234. ISSN: 0968-090X (page 16).

[108] X. Wang et al. "Implementation of Variable Speed Limits: Preliminary Test on Whitemud Drive, Edmonton, Canada". In: *Journal of Transportation Engineering* 142.12 (2016), p. 05016007 (page 65).

[109] Y. Wang and P. A. Ioannou. "New Model for Variable Speed Limits". In: *Transportation Research Record* 2249.1 (2011), pp. 38–43 (page 65).

[110] Washington State Department of Transportation Active Traffic and Demand Management http://www.wsdot.wa.gov/travel/operations-services/active-traffic-management/home. accessed 1-May-2019 (page 65).

[111] B. Williams and L. Hoel. "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results". In: *Journal of Transportation Engineering* 129.6 (2003), pp. 664–672 (page 16).

[112] B. Wolshon and A. Pande. *Traffic Engineering Handbook, 7th Edition*. John Wiley and Sons Inc., 2016 (page 12).

[113] C.-J. Wu et al. "Microsimulation of Congested Freeway with Multiple Vehicle Classes Using AIMSUN: A Case Study on SR-99N". In: *Transportation Research Board, 95th Annual Meeting of the Transportation Research Board, Washington, D.C.* 16-2974. 2016 (pages 44, 45).

[114] H. Zackor. "Speed Limitation on Freeways: Traffic-Responsive Strategies". In: *Concise Encyclopedia of Traffic & Transportation Systems*. Ed. by M. PAPAGEORGIOU. Amsterdam: Pergamon, 1991, pp. 507 –511. ISBN: 978-0-08-036203-8 (page 65).

[115] H. Zhang and W. Recker. "On optimal freeway ramp control policies for congested traffic corridors". In: *Transportation Research Part B: Methodological* 33.6 (1999), pp. 417 –436. ISSN: 0191-2615 (page 12).

[116] L. Zhang and D. Levinson. "Ramp metering and freeway bottleneck capacity". In: *Transportation Research Part A: Policy and Practice* 44.4 (2010), pp. 218 –235. ISSN: 0965-8564 (page 12).

[117] Y. Zhang and Y. Xie. "Forecasting of short-term freeway volume with v-support vector machines". In: *Transportation Research Record: Journal of the Transportation Research Board* 2024.1 (2008), pp. 92–99 (page 16).

[118] X. Zhou and H. S. Mahmassani. "A structural state space model for real-time traffic origindestination demand estimation and prediction in a day-to-day learning framework". In: *Transportation Research Part B: Methodological* 41.8 (2007), pp. 823 –840. ISSN: 0191-2615 (page 15).

Appendix A

Monitoring of Queue Length by Google Map on SR99-N Test Site

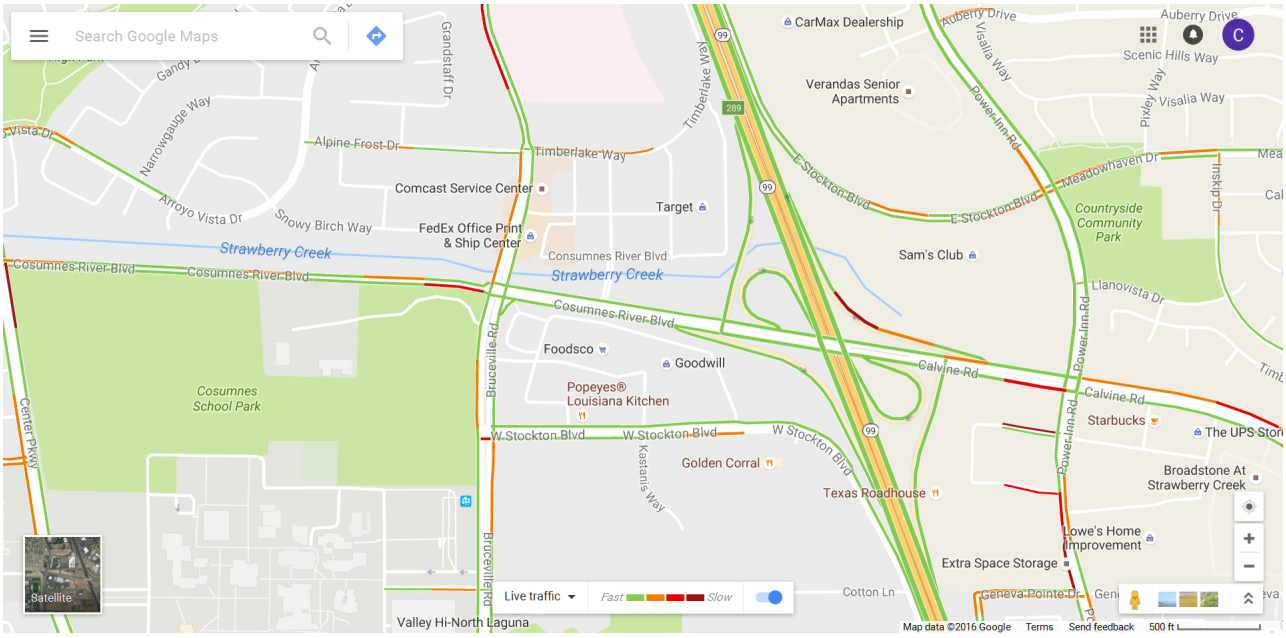


Figure A.1: Monitoring of queue length near EB Calvine Road on-ramp (OR6) and WB Calvine Road on-ramp (OR 7) at 8:02 AM on Wednesday, 12 October, 2016

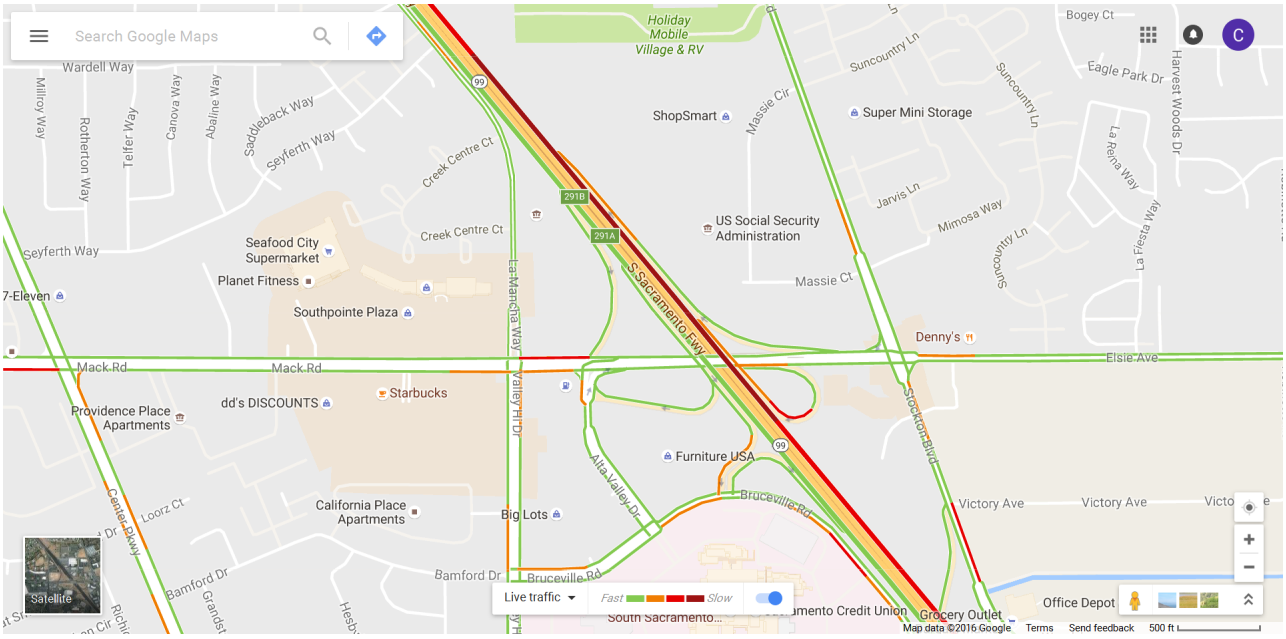


Figure A.2: Monitoring of queue length near EB Mack Road on-ramp (OR 8) and WB Mack Road on-ramp (OR 9) at 8:02 AM on Wednesday, 12 October, 2016

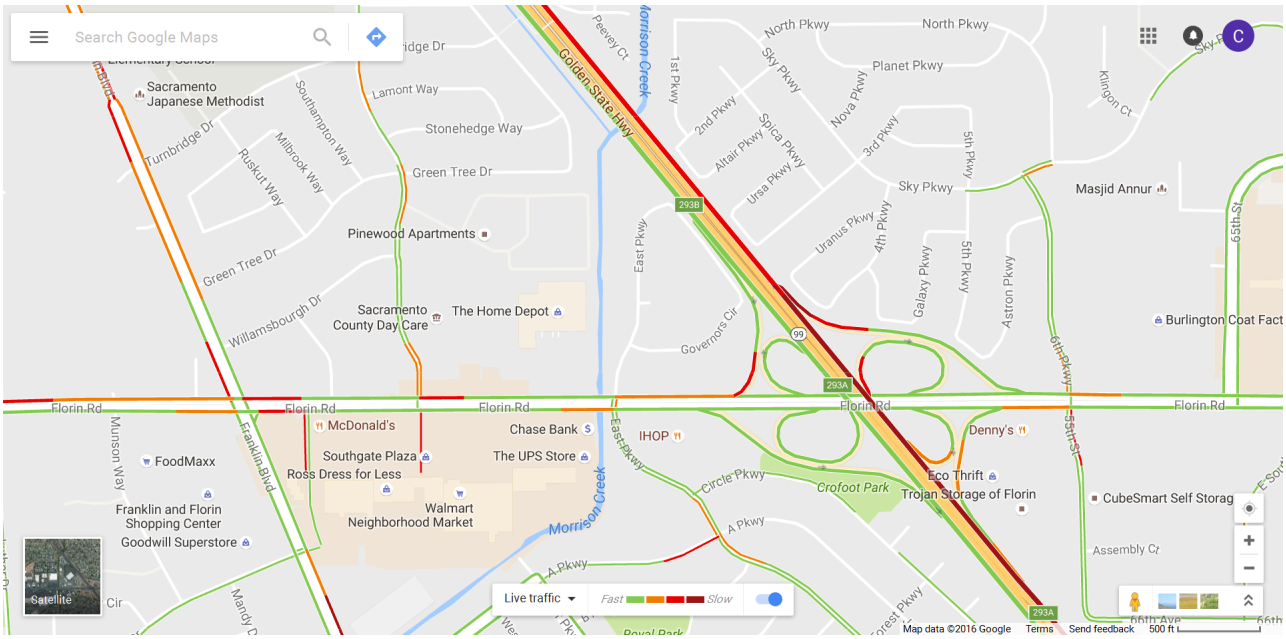


Figure A.3: Monitoring of queue length near EB Florin Road on-ramp (OR 10) and WB Florin Road on-ramp (OR 11) at 8:02 AM on Wednesday, 12 October, 2016

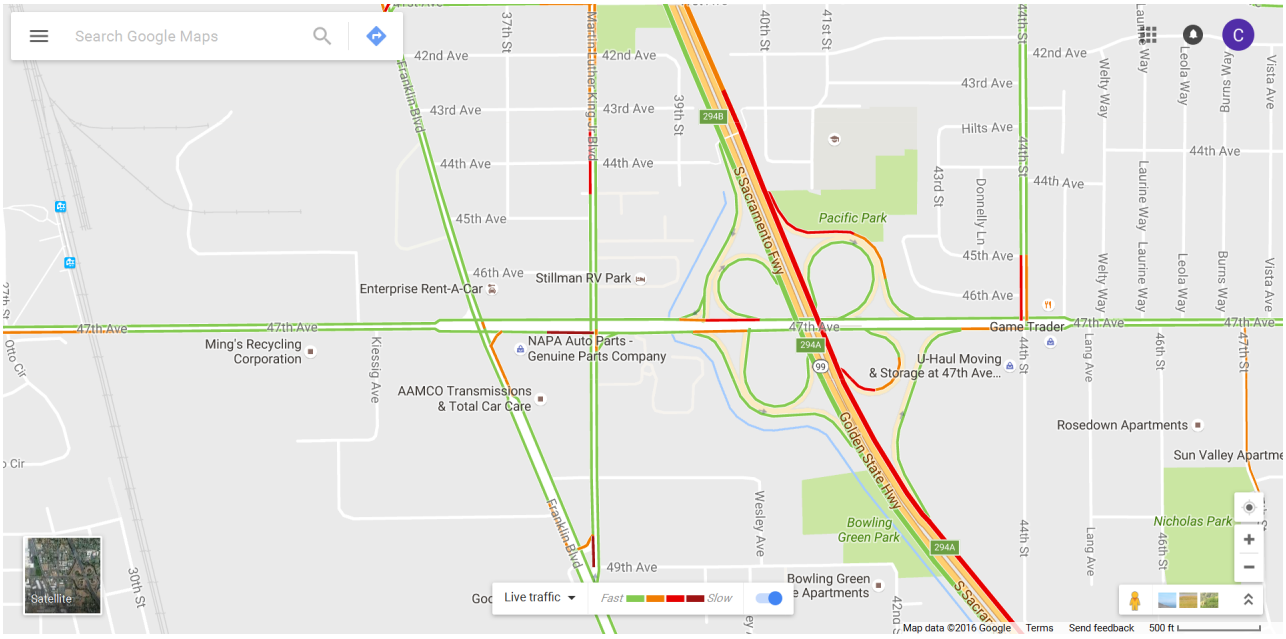


Figure A.4: Monitoring of queue length near EB 47th Ave on-ramp (OR 12) and WB 47th Ave on-ramp (OR 13) at 8:02 AM on Wednesday, 12 October, 2016

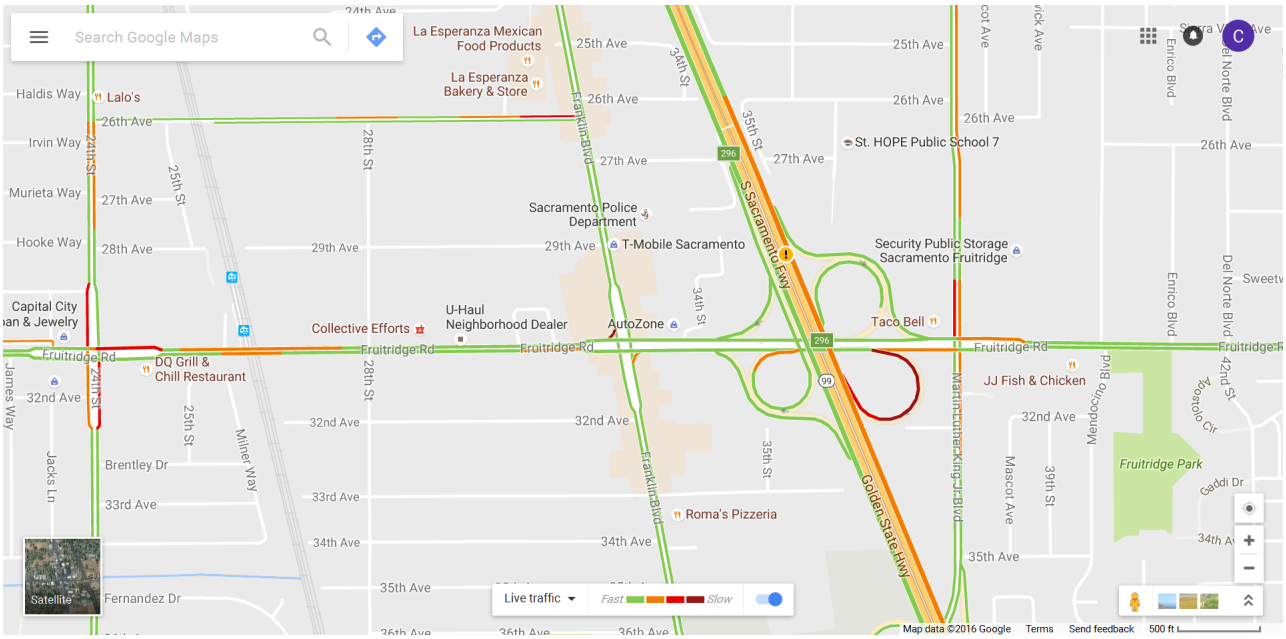


Figure A.5: Monitoring of queue length near EB Fruitridge Road on-ramp (OR 14) and WB Fruitridge Road on-ramp (OR 15) at 8:02 AM on Wednesday, 12 October, 2016

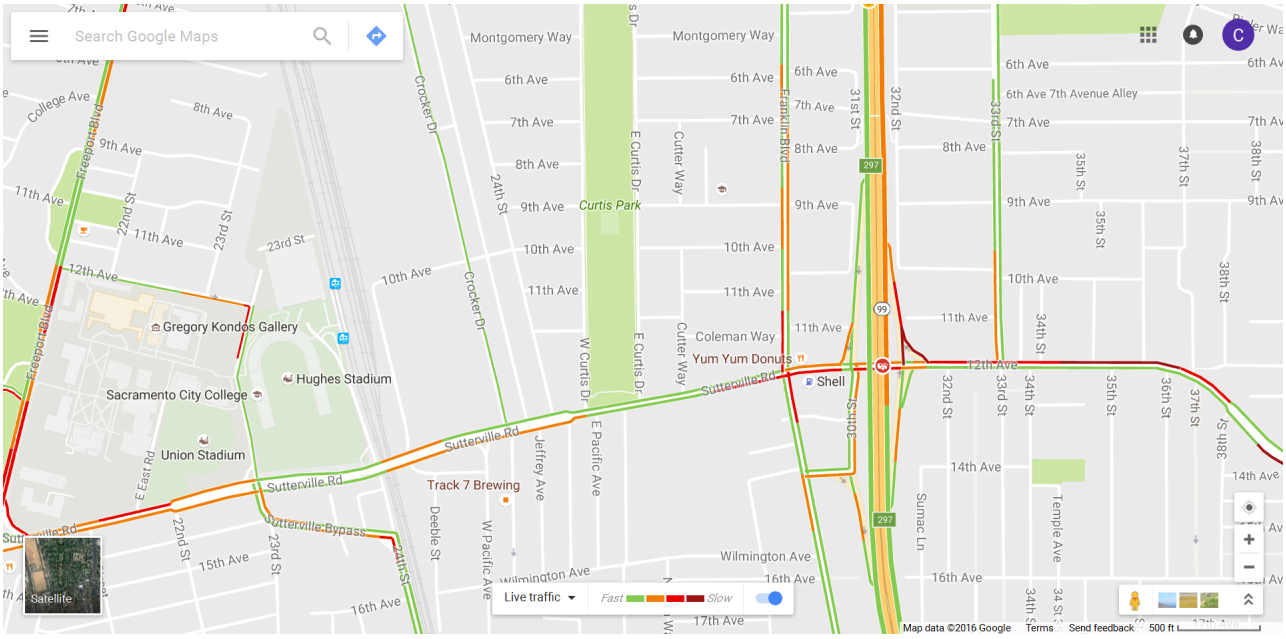


Figure A.6: Monitoring of queue length near 12th Ave on-ramp (OR 16) at 8:02 AM on Wednesday, 12 October, 2016

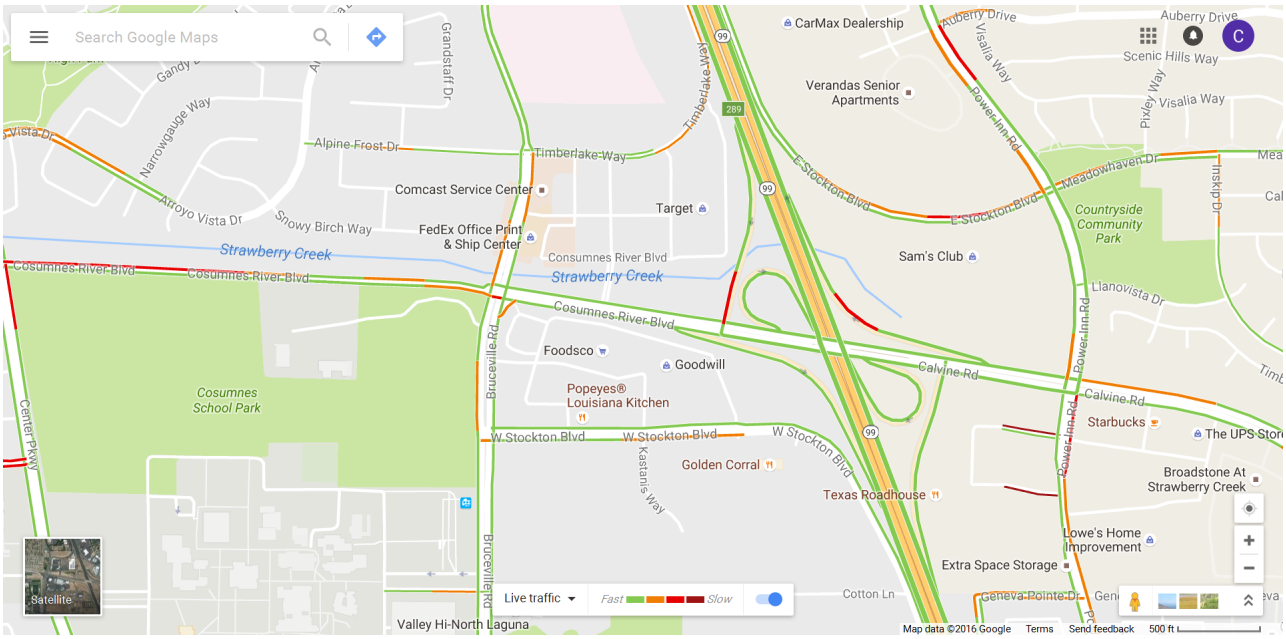


Figure A.7: Monitoring of queue length near EB Calvine Road on-ramp (OR6) and WB Calvine Road on-ramp (OR 7) at 8:22 AM on Wednesday, 12 October, 2016

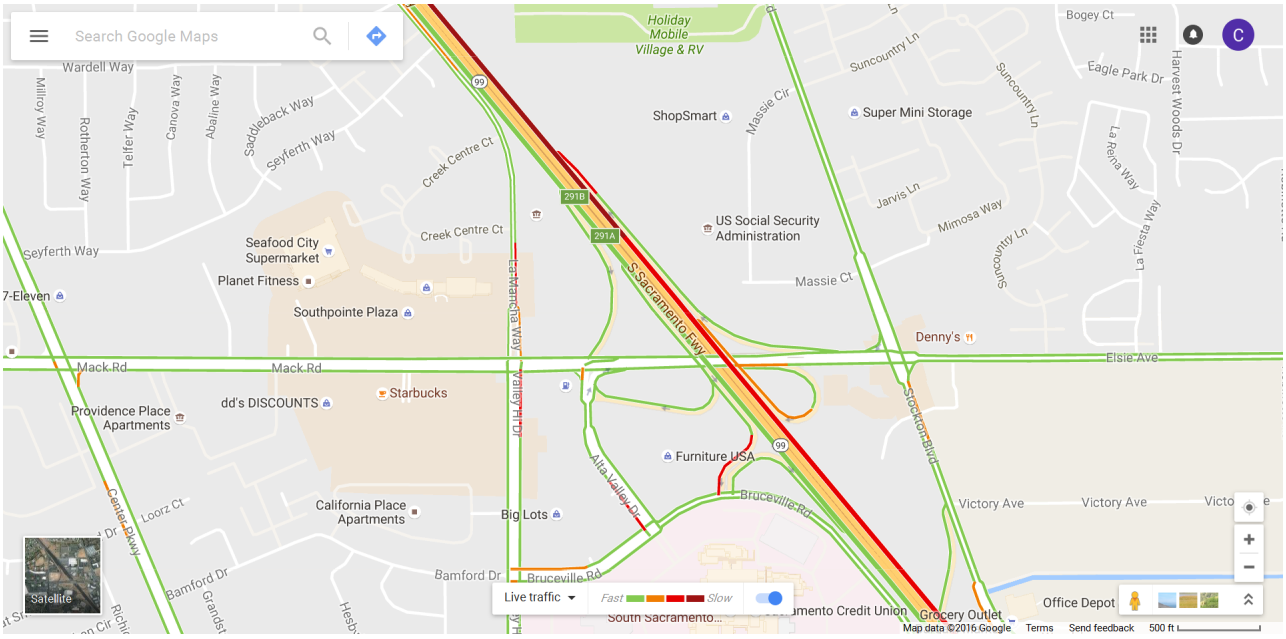


Figure A.8: Monitoring of queue length near EB Mack Road on-ramp (OR 8) and WB Mack Road on-ramp (OR 9) at 8:22 AM on Wednesday, 12 October, 2016

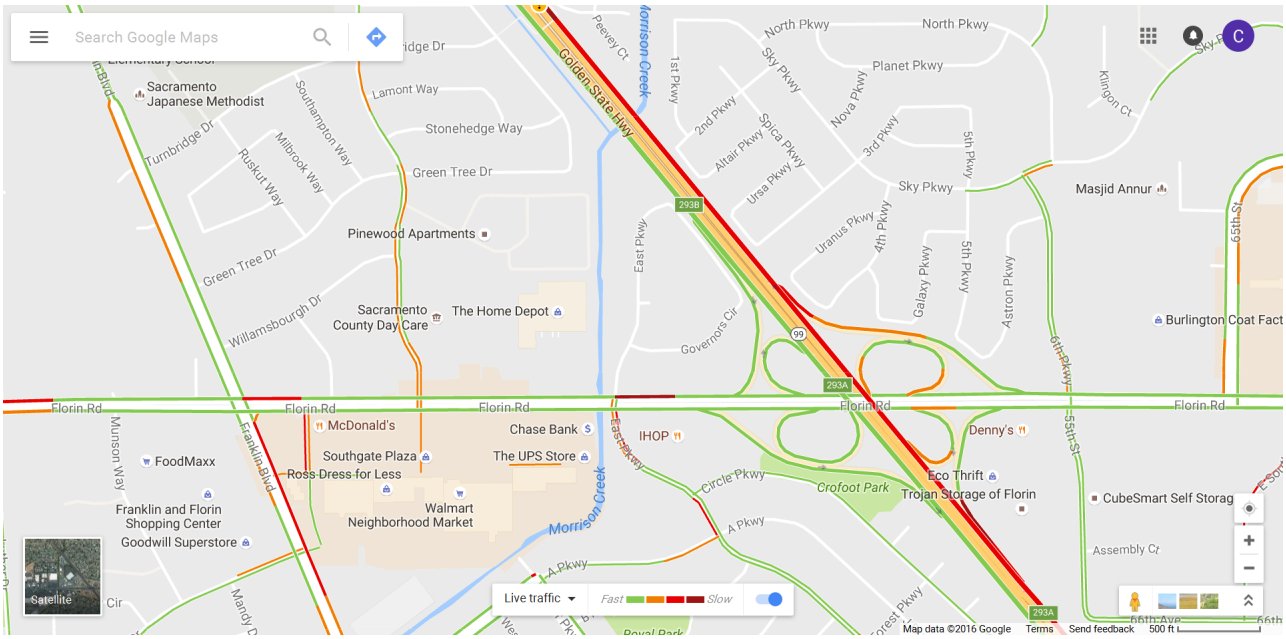


Figure A.9: Monitoring of queue length near EB Florin Road on-ramp (OR 10) and WB Florin Road on-ramp (OR 11) at 8:22 AM on Wednesday, 12 October, 2016

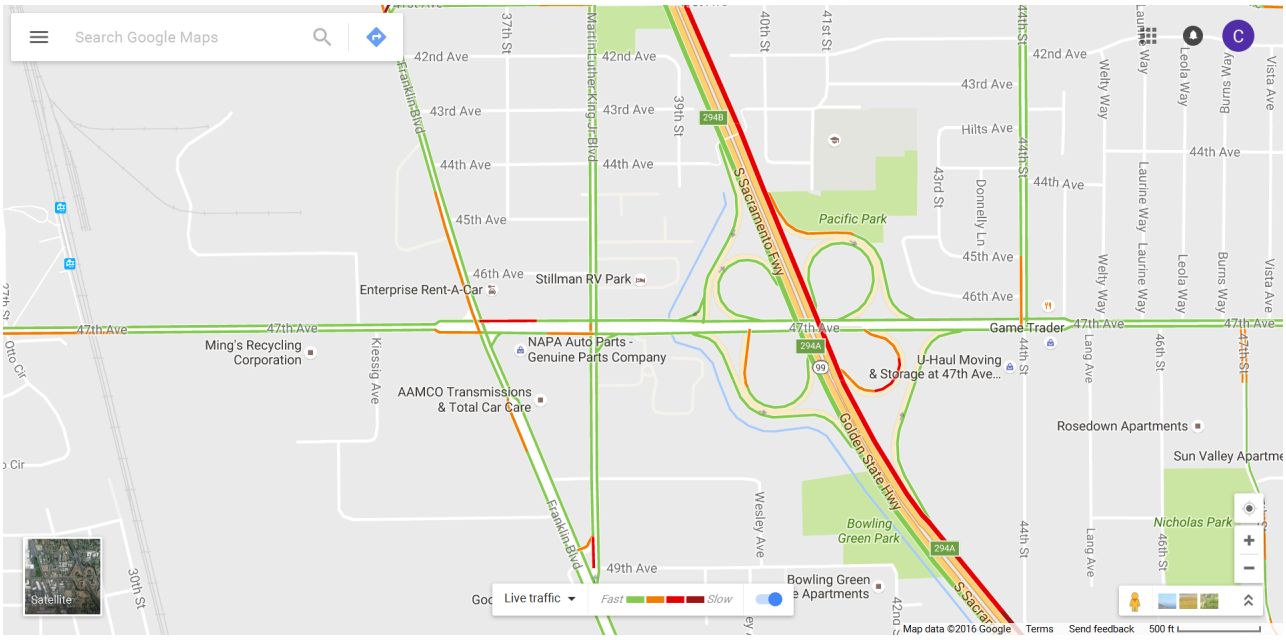


Figure A.10: Monitoring of queue length near EB 47th Ave on-ramp (OR 12) and WB 47th Ave on-ramp (OR 13) at 8:22 AM on Wednesday, 12 October, 2016

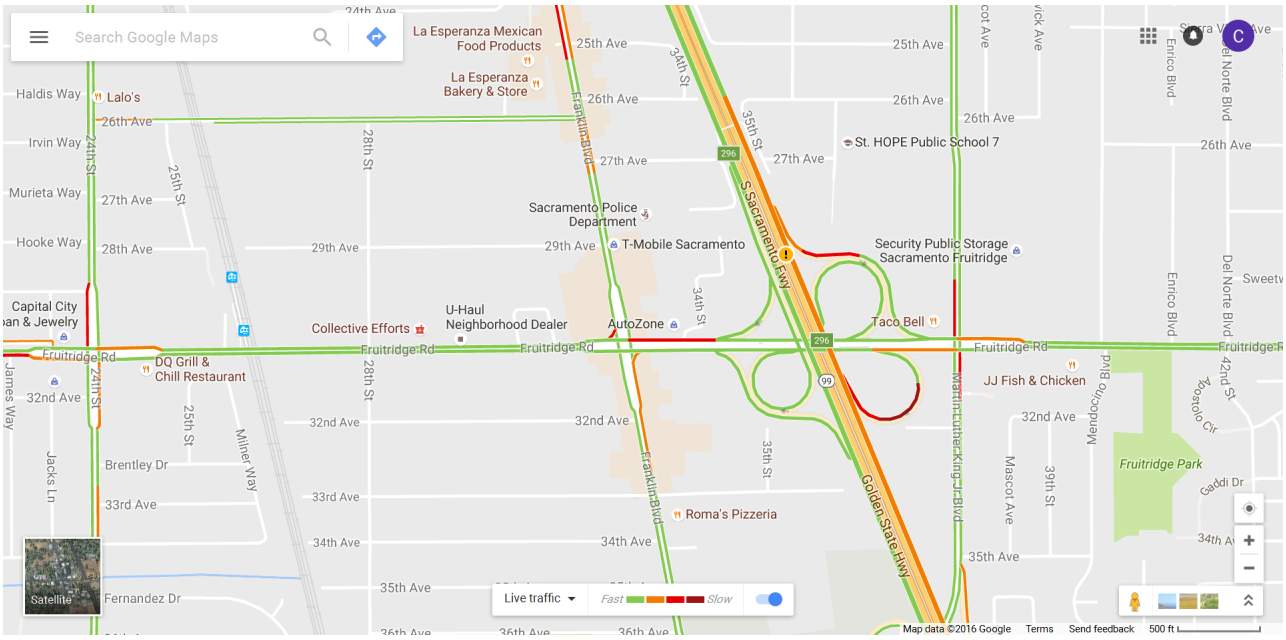


Figure A.11: Monitoring of queue length near EB Fruitridge Road on-ramp (OR 14) and WB Fruitridge Road on-ramp (OR 15) at 8:22 AM on Wednesday, 12 October, 2016

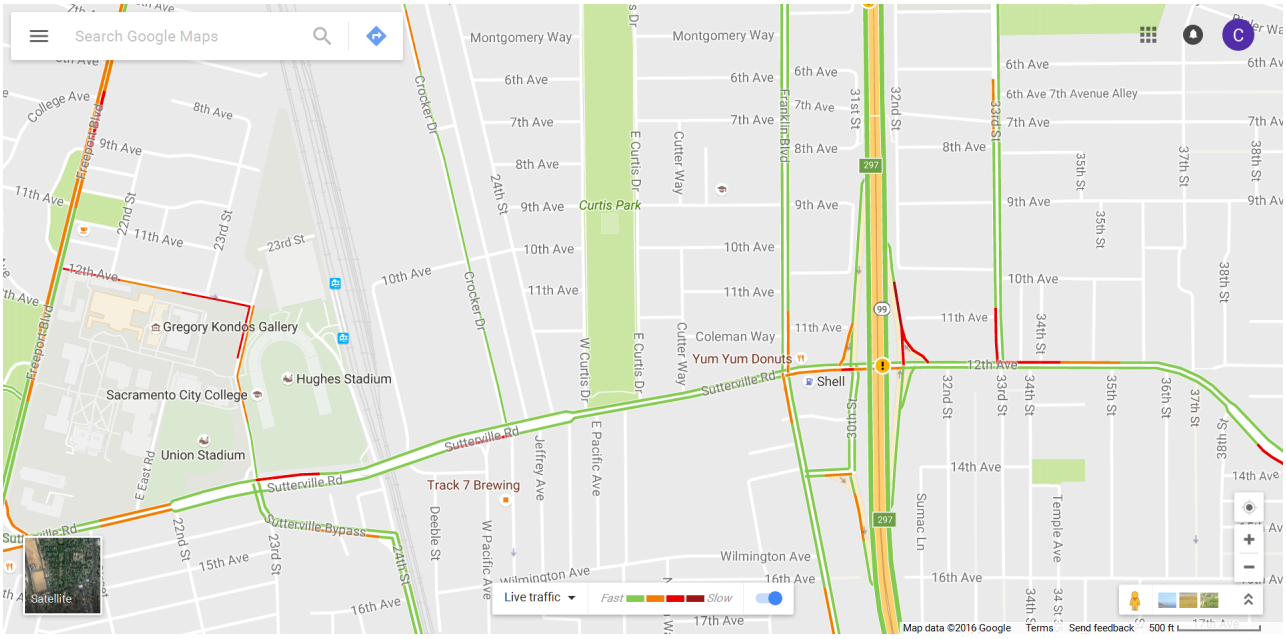


Figure A.12: Monitoring of queue length near 12th Ave on-ramp (OR 16) at 8:22 AM on Wednesday, 12 October, 2016

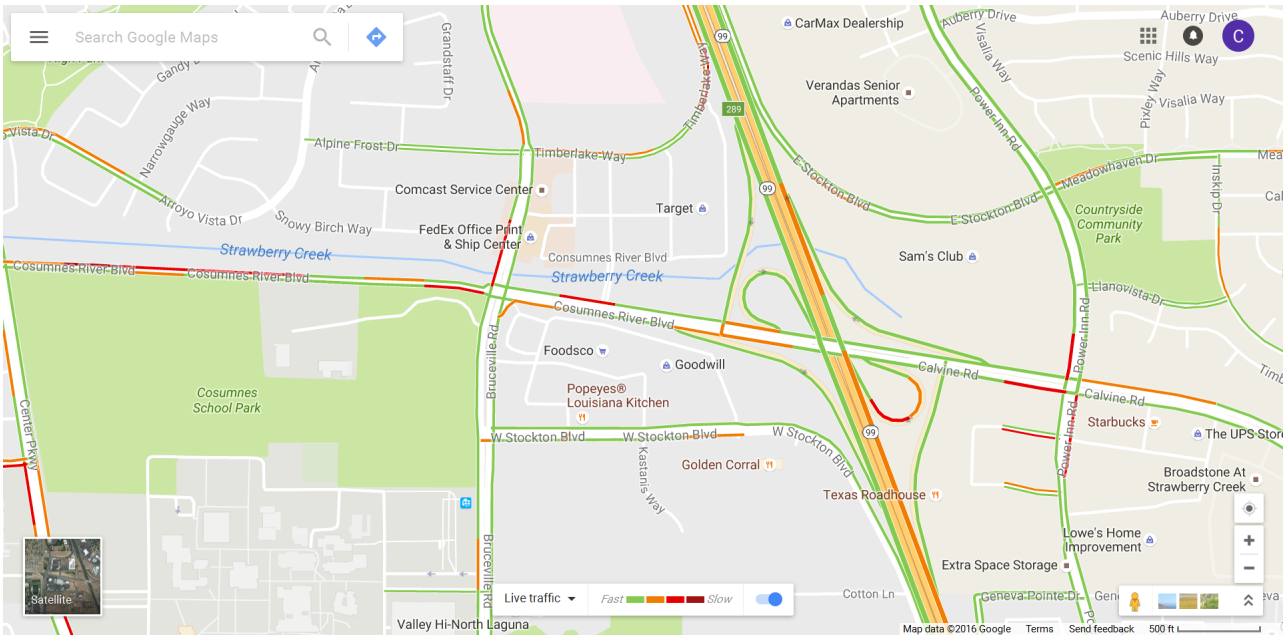


Figure A.13: Monitoring of queue length near EB Calvine Road on-ramp (OR6) and WB Calvine Road on-ramp (OR 7) at 9:03 AM on Wednesday, 12 October, 2016

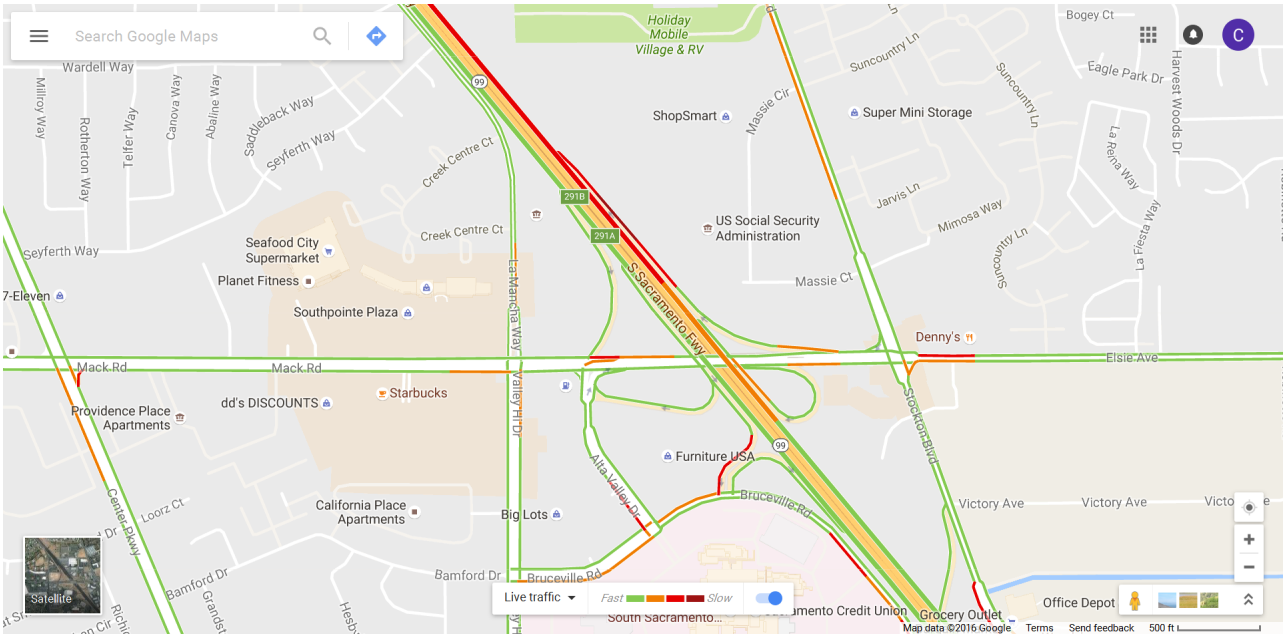


Figure A.14: Monitoring of queue length near EB Mack Road on-ramp (OR 8) and WB Mack Road on-ramp (OR 9) at 9:03 AM on Wednesday, 12 October, 2016

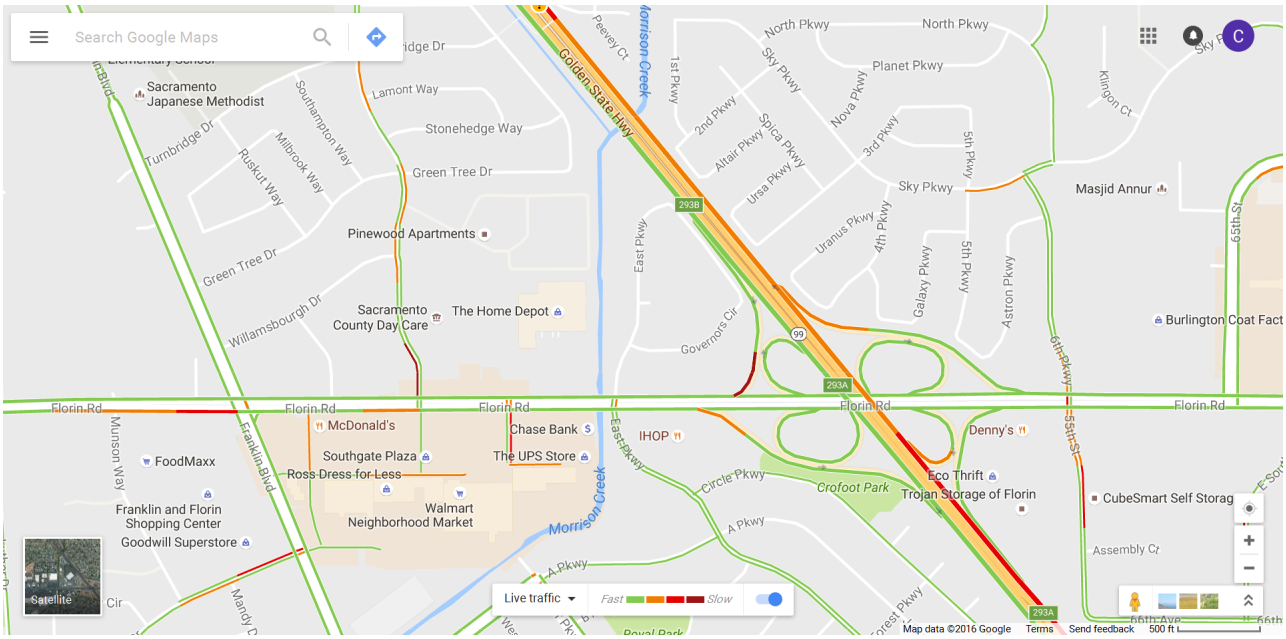


Figure A.15: Monitoring of queue length near EB Florin Road on-ramp (OR 10) and WB Florin Road on-ramp (OR 11) at 9:03 AM on Wednesday, 12 October, 2016

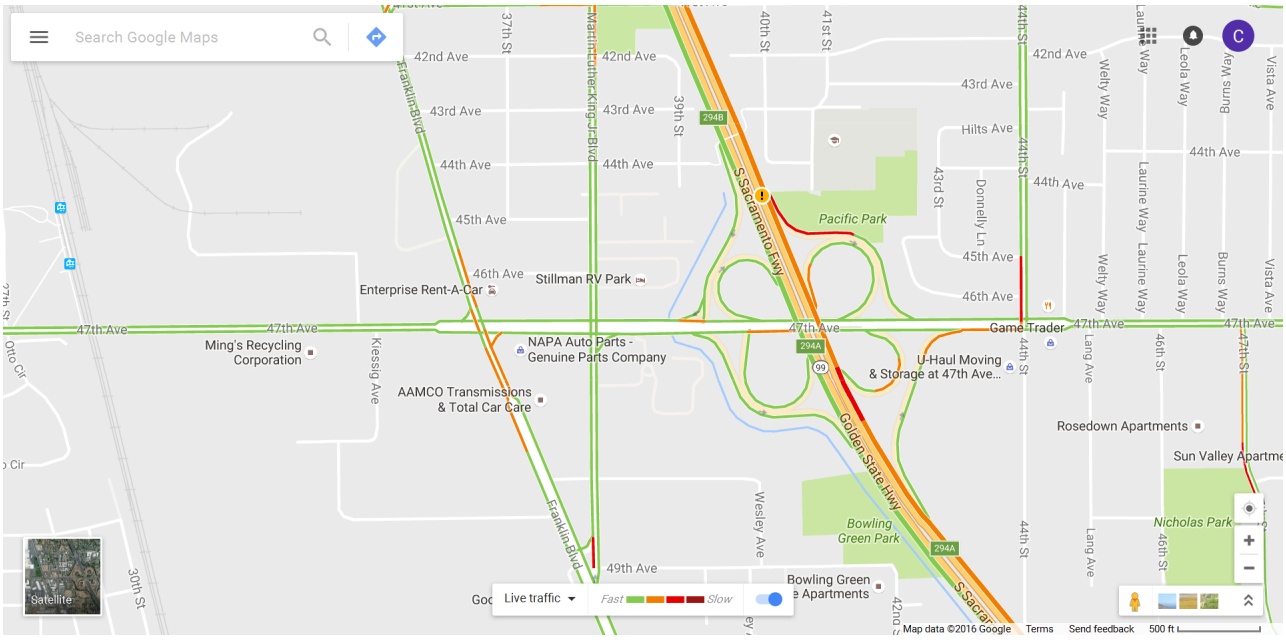


Figure A.16: Monitoring of queue length near EB 47th Ave on-ramp (OR 12) and WB 47th Ave on-ramp (OR 13) at 9:03 AM on Wednesday, 12 October, 2016

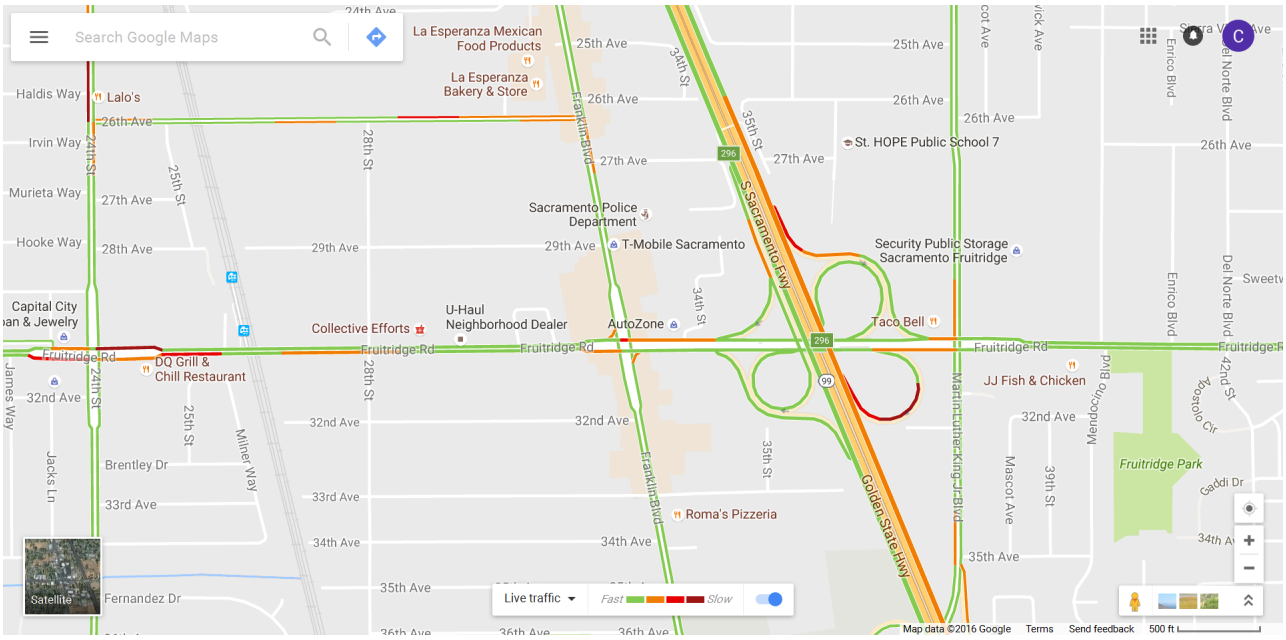


Figure A.17: Monitoring of queue length near EB Fruitridge Road on-ramp (OR 14) and WB Fruitridge Road on-ramp (OR 15) at 9:03 AM on Wednesday, 12 October, 2016

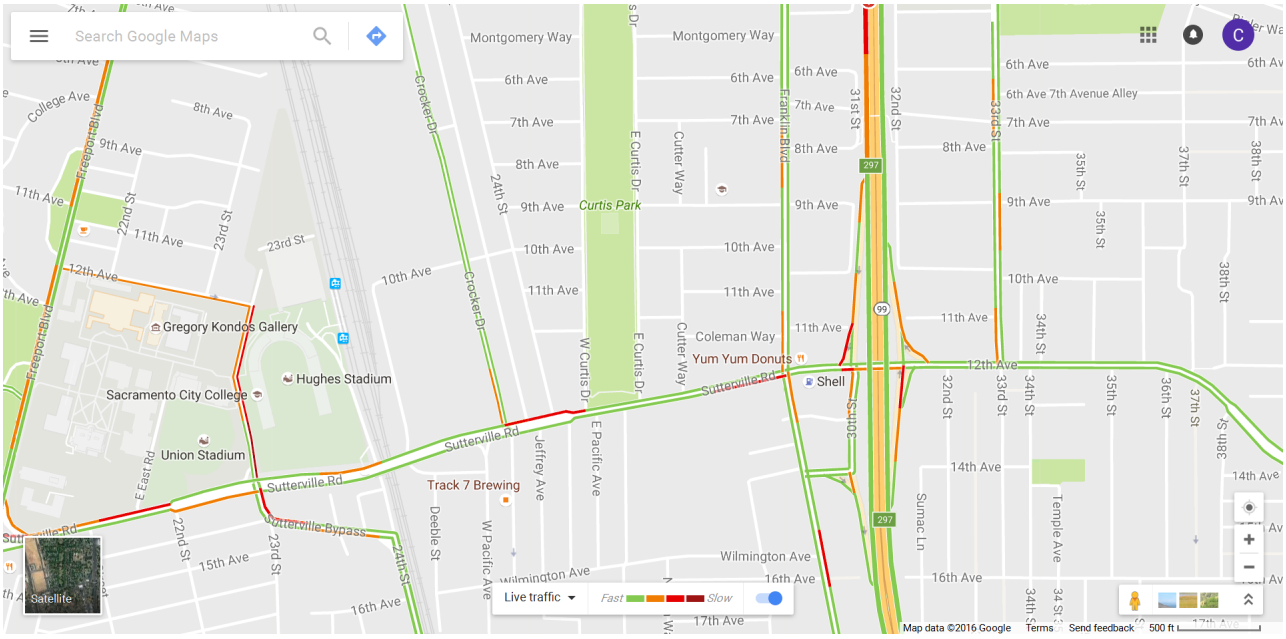


Figure A.18: Monitoring of queue length near 12th Ave on-ramp (OR 16) at 9:03 AM on Wednesday, 12 October, 2016